

09/475704

146 938

STIC-Biotech/ChemLib

From: Whiteman, Brian  
Sent: Monday, March 07, 2005 7:20 AM  
To: STIC-Biotech/ChemLib  
Subject: seq search

09/475,704 Barnett et al. 12/30/99

SEQ ID NO: 3 and 4

Please search protein and NA databases

- 1) default
- 2) interference

Thank you,

Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

RECEIVED  
MAR -7 2005  
STIC

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence :# \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIS: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 21:18:31 ; Search time 916.106 Seconds  
(without alignments)  
9798.345 Million cell updates/sec

Title: US-09-475-704A-4  
Perfect score: 1509

Sequence: 1 atggcgcccgccgcagcat.....gcggcccccctgagccagtaa 1509

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5537552 seqs, 2974263231 residues

Total number of hits satisfying chosen parameters: 11075104

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1509	100.0	1509	10	US-09-899-575-4 Sequence 4, Appli
2	1507.4	99.9	1509	10	US-09-967-464-64 Sequence 64, Appli
3	1504.2	99.7	1509	10	US-09-967-464-68 Sequence 68, Appli
4	1489.8	98.7	1509	10	US-09-899-575-21 Sequence 21, Appli
5	1297.6	86.0	1494	10	US-09-899-575-51 Sequence 51, Appli
6	1294.8	85.8	1491	10	US-09-899-575-99 Sequence 99, Appli
7	1276.8	84.6	1479	10	US-09-967-464-63 Sequence 63, Appli
8	1276.8	84.6	1479	10	US-09-899-575-3 Sequence 3, Appli
9	1275.2	84.5	3162	15	US-10-190-435-18 Sequence 18, Appli
10	1275.2	84.5	3462	15	US-10-190-435-16 Sequence 16, Appli
11	1275.2	84.5	4419	15	US-10-190-435-19 Sequence 19, Appli

12	1275.2	84.5	4419	16	US-10-190-305A-14	Sequence 14, Appli
13	1275.2	84.5	4483	15	US-10-190-435-35	Sequence 35, Appli
14	1275.2	84.5	4606	15	US-10-190-435-34	Sequence 34, Appli
15	1275.2	84.5	4615	15	US-10-190-435-36	Sequence 36, Appli
16	1275.2	84.5	4702	15	US-10-190-435-38	Sequence 38, Appli
17	1275.2	84.5	4716	15	US-10-190-435-17	Sequence 17, Appli
18	1275.2	84.5	4716	16	US-10-190-305A-13	Sequence 13, Appli
19	1273.6	84.4	2742	15	US-10-190-435-20	Sequence 20, Appli
20	1273.6	84.4	2742	16	US-10-190-305A-15	Sequence 15, Appli
21	1273.6	84.4	3930	15	US-10-190-435-9	Sequence 9, Appli
22	1273.6	84.4	3930	15	US-10-190-435-10	Sequence 10, Appli
23	1273.6	84.4	3930	15	US-10-190-435-11	Sequence 11, Appli
24	1273.6	84.4	5145	15	US-10-190-435-12	Sequence 12, Appli
25	1273.6	84.4	5145	16	US-10-190-305A-12	Sequence 12, Appli
26	1272.6	84.3	4713	15	US-10-190-435-59	Sequence 59, Appli
27	1272.6	84.3	4713	16	US-10-190-305A-83	Sequence 83, Appli
28	1271.6	84.3	5184	15	US-10-190-435-58	Sequence 58, Appli
29	1271.6	84.3	5184	16	US-10-190-305A-82	Sequence 82, Appli
30	1271.2	84.2	1479	10	US-09-967-464-67	Sequence 67, Appli
31	1270.2	84.2	2742	15	US-10-190-435-57	Sequence 57, Appli
32	1270.2	84.2	2742	16	US-10-190-305A-81	Sequence 81, Appli
33	1260.8	83.6	1479	10	US-09-899-575-20	Sequence 20, Appli
34	1201	79.6	1515	17	US-10-387-336-4	Sequence 4, Appli
35	1201	79.6	4472	17	US-10-387-336-75	Sequence 75, Appli
36	1201	79.6	4608	17	US-10-387-336-76	Sequence 76, Appli
37	1201	79.6	4689	17	US-10-387-336-74	Sequence 74, Appli
38	1201	79.6	4766	17	US-10-387-336-73	Sequence 73, Appli
39	1199.4	79.5	2799	16	US-10-241-009-18	Sequence 18, Appli
40	1199.4	79.5	2799	16	US-10-190-434B-18	Sequence 18, Appli
41	1199.4	79.5	2799	16	US-10-190-305A-11	Sequence 11, Appli
42	1199.4	79.5	3205	16	US-10-241-009-17	Sequence 17, Appli
43	1199.4	79.5	3205	16	US-10-190-434B-17	Sequence 17, Appli
44	1199.4	79.5	3496	16	US-10-241-009-15	Sequence 15, Appli
45	1199.4	79.5	3496	16	US-10-190-434B-15	Sequence 15, Appli

## ALIGNMENTS

RESULT 1  
US-09-899-575-4  
; Sequence 4, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megele, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; PRIOR FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 09/475,704  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1509  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag  
; OTHER INFORMATION: of HIV strain AF110967  
US-09-899-575-4

Query Match 100.0%; Score 1509; DB 10; Length 1509;  
Best local Similarity 100.0%; Pred. No. 0;  
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCCCGCCGCGCAGCATCTCTGCGCGCGAGAACTGACAACTGGAGAGATCCGC 60  
Db 1 ATGGCGCCCGCCGCGCAGCATCTCTGCGCGCGAGAACTGACAACTGGAGAGATCCGC 60

```
QY 61 CTGCCCCCGCGGCAAGACACTATGCTGAAGCACTGTGTGGCCAGCCCGGAG 120
Db 61 CTGCCCCCGCGGCAAGACACTATGCTGAAGCACTGTGTGGCCAGCCCGGAG 120
QY 121 CTGAGGGGCTTCCGCTGAACCCCGGCTGTGAGACCCGCGGAGATC 180
Db 121 CTGAGGGGCTTCCGCTGAACCCCGGCTGTGAGACCCGCGGAGATC 180
QY 181 ATGAAGCAGCTGCAGCCCCCGCTGCAGACCCGAGAGCTGGCGACCTGTACAAC 240
Db 181 ATGAAGCAGCTGCAGCCCCCGCTGCAGACCCGAGAGCTGGCGACCTGTACAAC 240
QY 241 ACCGTGGCCACCTGTACTGCGTGCAAGCCGCGCATGAGTCCGCGACACCAAGAGGCC 300
Db 241 ACCGTGGCCACCTGTACTGCGTGCAAGCCGCGCATGAGTCCGCGACACCAAGAGGCC 300
QY 301 CTGGAACAAGATCGAGAGAGAGACAGAACTAAGTCCCAAGAGAGAGACCCAGCGCAAGAG 360
Db 301 CTGGAACAAGATCGAGAGAGAGACAGAACTAAGTCCCAAGAGAGAGACCCAGCGCAAGAG 360
QY 361 GCGGACGGCAAGGTGAGCCAGAACTAAGTCCGCGCATGAGTCCGCGAGAGAGAGAGAG 420
Db 361 GCGGACGGCAAGGTGAGCCAGAACTAAGTCCGCGCATGAGTCCGCGAGAGAGAGAGAG 420
QY 421 CACCAAGGCGCATCAGCCCCCGGCAAGCTGAAGCGCTGGGTGAAGGTGATCGAGAGAGAG 480
Db 421 CACCAAGGCGCATCAGCCCCCGGCAAGCTGAAGCGCTGGGTGAAGGTGATCGAGAGAGAG 480
QY 481 TTCAAGCCCCGAGGTGATCCCCCATTTCAAGCGGCTGAGCGAGGGGCCACCCCGGAGAG 540
Db 481 TTCAAGCCCCGAGGTGATCCCCCATTTCAAGCGGCTGAGCGAGGGGCCACCCCGGAGAG 540
QY 541 CTGAACACGATGTTGAACACCGTGGGGCGGCAAGCGCGCGCATGAGTCTGAAGAGAC 600
Db 541 CTGAACACGATGTTGAACACCGTGGGGCGGCAAGCGCGCGCATGAGTCTGAAGAGAC 600
QY 601 ACCATCAACGAGAGAGCGCGCGGAGTGGGAGCCGCTGCAAGCCCGTGCAGAGCGCGCGCTG 660
Db 601 ACCATCAACGAGAGAGCGCGCGGAGTGGGAGCCGCTGCAAGCCCGTGCAGAGCGCGCGCTG 660
QY 661 GCGCGCGCGCGGAGATGCGGAGACCCCGCGGAGAGCATCGCGCGGCGCAACGAGACCTG 720
Db 661 GCGCGCGCGCGGAGATGCGGAGACCCCGCGGAGAGCATCGCGCGGCGCAACGAGACCTG 720
QY 721 CAGGAGCAGATCGCCTGGATGACAGCAACCCCGCGTGGCGCGGAGCATCTACAAG 780
Db 721 CAGGAGCAGATCGCCTGGATGACAGCAACCCCGCGTGGCGCGGAGCATCTACAAG 780
QY 781 CGGTGGATCATCTGGGGCTGAACAAGATCGTGGGATGTACAGCGCGCGTGAAGATCTCTG 840
Db 781 CGGTGGATCATCTGGGGCTGAACAAGATCGTGGGATGTACAGCGCGCGTGAAGATCTCTG 840
QY 841 GACATCCGCGGAGGCGCGGAGAGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900
Db 841 GACATCCGCGGAGGCGCGGAGAGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900
QY 901 CTGCGCGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Db 901 CTGCGCGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 961 CAGAAAGCGCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCGCGCGGCGCAACCTG 1020
Db 961 CAGAAAGCGCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCGCGCGGCGCAACCTG 1020
QY 1021 GAGGAGATGATGACCGCTGCGAGGGGGCTGGGGCGCGCGCGCGCAAGAGCGCGGTGCTG 1080
Db 1021 GAGGAGATGATGACCGCTGCGAGGGGGCTGGGGCGCGCGCGCGCAAGAGCGCGGTGCTG 1080
QY 1081 GCGGAGGCGATGAGCGGCGCAAGCGGTGAACATCATGATGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 GCGGAGGCGATGAGCGGCGCAAGCGGTGAACATCATGATGAGAGAGAGAGAGAGAGAGAG 1140
```

```
QY 1141 GCGCGCGCGGCGCAAGCTCAAGTGTCTCAACTGCGGCAAGAGGCGCACATCGCGCAAGAG 1200
Db 1141 GCGCGCGCGGCGCAAGCTCAAGTGTCTCAACTGCGGCAAGAGGCGCACATCGCGCAAGAG 1200
QY 1201 TGCGCGCGCGCGCGCAAGAGAGGCTGTGAAAGTGCAGAGAGAGGCGCACAGATGAAG 1260
Db 1201 TGCGCGCGCGCGCGCAAGAGAGGCTGTGAAAGTGCAGAGAGAGGCGCACAGATGAAG 1260
QY 1261 GACTGACCGAGCGCGCAGGCGCAACTTCTGCGGCAAGATCTGCGGCGGCGCACAGAGGCG 1320
Db 1261 GACTGACCGAGCGCGCAGGCGCAACTTCTGCGGCAAGATCTGCGGCGGCGCACAGAGGCG 1320
QY 1321 CCGGCAACTTCTGCGAAGACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 1321 CCGGCAACTTCTGCGAAGACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
QY 1381 CCGGCGGAGAGCTTCGCTGCGAAGACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCG 1440
Db 1381 CCGGCGGAGAGCTTCGCTGCGAAGACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCG 1440
QY 1441 GCGGAGCGCTTCCGCTGCGAGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
Db 1441 GCGGAGCGCTTCCGCTGCGAGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
QY 1501 AGCCAGTAA 1509
Db 1501 AGCCAGTAA 1509
```

RESULT 2  
US-09-967-464-64

```
; Sequence 64, Application US/09967464  
; Publication No. US20030138453A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Hagan, Derek  
; APPLICANT: Otten, Gillis  
; APPLICANT: Donnelly, John J.  
; APPLICANT: Polo, John M.  
; APPLICANT: Barnett, Susan  
; APPLICANT: Singh, Mamohan  
; APPLICANT: Ulmer, Jeffrey  
; APPLICANT: Dubensky, Jr., Thomas W.  
; TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF HETEROLOGOUS NUCLEIC ACIDS  
; FILE REFERENCE: P16269.004  
; CURRENT APPLICATION NUMBER: US/09/967,464  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/236,105  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: 60/315,905  
; PRIOR FILING DATE: 2001-08-30  
; NUMBER OF SEQ ID NOS: 68  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64  
; LENGTH: 1509  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-967-464-64
```

```
Query Match 99.9%; Score 1507.4; DB 10; Length 1509;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1508; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGGGCGCGCGCGCAGACATCTGCGCGGCGAGAGAGTGGAGAAAGATCCGC 60  
Db 1 ATGGGCGCGCGCGCAGACATCTGCGCGGCGAGAGAGTGGAGAAAGATCCGC 60  
QY 61 CTGCGCGCGCGCGCAAGAGACTACATGCTGAAGCACTGTGTGGCCAGCGCGAG 120  
Db 61 CTGCGCGCGCGCGCAAGAGACTACATGCTGAAGCACTGTGTGGCCAGCGCGAG 120  
QY 121 CTGAGGGGCTTCCGCTGAACCCCGGCTGTGAGAGCGCGGAGGCTGCAAGAGATC 180  
Db 121 CTGAGGGGCTTCCGCTGAACCCCGGCTGTGAGAGCGCGGAGGCTGCAAGAGATC 180
```

```
QY 181 ATGAGCAGCTGACGCCCCCTGCAACCGGACCGAGGAGCTGGCCAGCTGTACAAC 240
Db 181 ATGAAGCAGCTGACGCCCCCTGCAACCGGACCGAGGAGCTGGCCAGCTGTACAAC 240
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGCCGGCATGAGGTCCGCGACACCAAGAGGCC 300
Db 241 ACCGTGGCCACCTGTACTGCGTGCACGCCGGCATGAGGTCCGCGACACCAAGAGGCC 300
QY 301 CTGGAACAAGATCGAGAGAGAGCAGAACAGTCCACAGAGAAGACCCAGCGCAAGAG 360
Db 301 CTGGAACAAGATCGAGAGAGAGCAGAACAGTCCACAGAGAAGACCCAGCGCAAGAG 360
QY 361 GCCGACGGCAAGGTGAGCCAGAACTAACCCTATGTCAGAACTGACAGGGCCAGATGGTG 420
Db 361 GCCGACGGCAAGGTGAGCCAGAACTAACCCTATGTCAGAACTGACAGGGCCAGATGGTG 420
QY 421 CACGAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGTGAATCGAGAGAGGCC 480
Db 421 CACGAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGTGAATCGAGAGAGGCC 480
QY 481 TTGAGCCCCGAGGTGATCCCCATGTTCAACCGCCCTGAGCGAGGGCCGACCCCGCAGAC 540
Db 481 TTGAGCCCCGAGGTGATCCCCATGTTCAACCGCCCTGAGCGAGGGCCGACCCCGCAGAC 540
QY 541 CTGAACACGATGTTGAACACCGTGGGCGGCGCACCGAGCCCGCATGACATGCTGAAGAC 600
Db 541 CTGAACACGATGTTGAACACCGTGGGCGGCGCACCGAGCCCGCATGACATGCTGAAGAC 600
QY 601 ACCATCAACGAGAGGCGCCCGCAGTGGGACCGCCTGCAACCCCGTGAAGGCGGCGCGTG 660
Db 601 ACCATCAACGAGAGGCGCCCGCAGTGGGACCGCCTGCAACCCCGTGAAGGCGGCGCGTG 660
QY 661 GCGCGCGCGCAGATGCGCGGACCCCGCGCGCAGCGACATCGCCGCGCCACCAAGCCTTG 720
Db 661 GCGCGCGCGCAGATGCGCGGACCCCGCGCGCAGCGACATCGCCGCGCCACCAAGCCTTG 720
QY 721 CAGGAGCAGATCGCCTGGATGACCAAGCAACCCCGCTGCGCGGCGACATCTACAAG 780
Db 721 CAGGAGCAGATCGCCTGGATGACCAAGCAACCCCGCTGCGCGGCGACATCTACAAG 780
QY 781 CCGTGGATCATCTGGGGCTGAAACAAGATCGTGGGATGTACAGCCCGCTGAGCATCTTG 840
Db 781 CCGTGGATCATCTGGGGCTGAAACAAGATCGTGGGATGTACAGCCCGCTGAGCATCTTG 840
QY 841 GACATCCGCGCAGGCGCCCAAGGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900
Db 841 GACATCCGCGCAGGCGCCCAAGGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900
QY 901 CTGCGCGCGCAGAGCAGGCCACCCAGAGCGTGAAGAACTGGATGACGAGACCTTGTGGTG 960
Db 901 CTGCGCGCGCAGAGCAGGCCACCCAGAGCGTGAAGAACTGGATGACGAGACCTTGTGGTG 960
QY 961 CAGAAAGCCAAACCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCGCACCTTG 1020
Db 961 CAGAAAGCCAAACCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCGCACCTTG 1020
QY 1021 GAGGAGATGATGACCGCTGCGCAGGGCGTGGGCGCGCGCGCACAAAGGCCGCTGCTG 1080
Db 1021 GAGGAGATGATGACCGCTGCGCAGGGCGTGGGCGCGCGCGCACAAAGGCCGCTGCTG 1080
QY 1081 GCGGAGGCGATGAGCCAGGCCCAAGCGTGAACATCATGATGACAGAGAGCAACTTCAAG 1140
Db 1081 GCGGAGGCGATGAGCCAGGCCCAAGCGTGAACATCATGATGACAGAGAGCAACTTCAAG 1140
QY 1141 GCGCGCGCGCGCAACGTCAAGTGTCTTCAACTGCGCAAGAGGCGCACATCGCCAAAGAAC 1200
Db 1141 GCGCGCGCGCGCAACGTCAAGTGTCTTCAACTGCGCAAGAGGCGCACATCGCCAAAGAAC 1200
QY 1201 TCGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGGCGCAAGGAGGCCACCAAGATGAAG 1260
Db 1201 TCGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGGCGCAAGGAGGCCACCAAGATGAAG 1260
```

```
QY 1261 GACTGCAACCGAGCGCCCAAGCCAACTTCTTGGGCAAGATCTGGCCCAAGAGGCGCC 1320
Db 1261 GACTGCAACCGAGCGCCCAAGCCAACTTCTTGGGCAAGATCTGGCCCAAGAGGCGCC 1320
QY 1321 CCGGCAACTTCTTGGGCAAGACCGGAGCGAGCCCGCGCGCCCAAGAGGAGGCCCGCC 1380
Db 1321 CCGGCAACTTCTTGGGCAAGACCGGAGCGAGCCCGCGCGCCCAAGAGGAGGCCCGCC 1380
QY 1381 CCGGCGGAGAGCTTCCGCTTGGAGAGACCAACCCCGCGCCCAAGAGAGGCCCAAGAG 1440
Db 1381 CCGGCGGAGAGCTTCCGCTTGGAGAGACCAACCCCGCGCCCAAGAGAGGCCCAAGAG 1440
QY 1441 CGGAGCGCTTACCGCGAGCGCCCTGACCGCCCTGCGCAGCGCTGTTGGCGAGCGCCCTG 1500
Db 1441 CGGAGCGCTTACCGCGAGCGCCCTGACCGCCCTGCGCAGCGCTGTTGGCGAGCGCCCTG 1500
QY 1501 AGCCAGTAA 1509
Db 1501 AGCCAGTAA 1509
```

```
RESULT 3
US-09-967-464-68
; Sequence 68, Application US/09967464
; Publication No. US20030138453A1
; GENERAL INFORMATION:
; APPLICANT: O'Hagan, Derek
; APPLICANT: Otten, Gilles
; APPLICANT: Donnelly, John J.
; APPLICANT: Polo, John M.
; APPLICANT: Barnett, Susan
; APPLICANT: Singh, Mamohan
; APPLICANT: Ulmer, Jeffrey
; APPLICANT: Dubensky, Jr., Thomas W.
; TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF HETEROLOGOUS NUCLEIC ACIDS
; FILE REFERENCE: P16269.004
; CURRENT APPLICATION NUMBER: US/09/967,464
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/236,105
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/315,905
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
US-09-967-464-68
```

```
Query Match 99.7%; Score 1504.2; DB 10; Length 1509;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1497; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGCGGAGAGCTGACAAAGTGGAGAGATCCGC 60
Db 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGCGGAGAGCTGACAAAGTGGAGAGATCCGC 60
QY 61 CTGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACCTGTGTGGGCCAGCGCGAG 120
Db 61 CTGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACCTGTGTGGGCCAGCGCGAG 120
QY 121 CTGAGAGGCTTCCGCTTGAACCCCGCGCTGCTGAGAGACCGCCGAGGGCTGCAAGCATC 180
Db 121 CTGAGAGGCTTCCGCTTGAACCCCGCGCTGCTGAGAGACCGCCGAGGGCTGCAAGCATC 180
QY 181 ATGAAGCAGCTGACGCGCGCGCTGCAAGACCGGACCGAGAGCTGCGCAGCTGTACAAC 240
Db 181 ATGAAGCAGCTGACGCGCGCGCTGCAAGACCGGACCGAGAGCTGCGCAGCTGTACAAC 240
QY 241 ACCGTGGCCACCTGTACTGCGTGCAAGCGCGCATGAGGTCCGCGACACCAAGAGGCC 300
Db 241 ACCGTGGCCACCTGTACTGCGTGCAAGCGCGCATGAGGTSCGCGACACCAAGAGGCC 300
```

```
QY 301 CTGACAGATCGAGAGAGAGACAAGTCCACGACAGAGACCCCAAGAGAG 360
Db 301 CTGACAGATCGAGAGAGAGACAAGTCCACGAGAGAGAGAGAGAGAGAG 360
QY 361 GCCGAGCGCAAGGTGAGCCAGACTACCCCATCTGTGAGAACTGAGAGGCCAGATGTG 420
Db 361 GCCGAGCGCAAGGTGAGCCAGACTACCCCATCTGTGAGAACTGAGAGGCCAGATGTG 420
QY 421 CACGAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAGAGCC 480
Db 421 CACGAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAGAGCC 480
QY 481 TTCAAGCCCGAGGTGATCCCATATTTCAACGCTGAGCGAGAGGCCGCCACCCCGAGAG 540
Db 481 TTCAAGCCCGAGGTGATCCCATATTTCAACGCTGAGCGAGAGGCCGCCACCCCGAGAG 540
QY 541 CTGAACACGATGTTGAACACCGTGGGCGGCCACCAAGGCCCATGACATGCTGAAGAG 600
Db 541 CTGAACACGATGTTGAACACCGTGGGCGGCCACCAAGGCCCATGACATGCTGAAGAG 600
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGGACCGCTGACACCCCGTGAAGGCCGCCCGTG 660
Db 601 ACCATCAACGAGAGAGCGCGCGAGTGGGACCGCTGACACCCCGTGAAGGCCGCCCGTG 660
QY 661 GCCCGCGCGAGATGCGCGGACCCCGCGCGAGCAATCGCGCGCGCACCAAGACCTG 720
Db 661 GCCCGCGCGAGATGCGCGGACCCCGCGCGAGCAATCGCGCGCGCACCAAGACCTG 720
QY 721 CAGGAGCAGATCGCTGATGACCAAGCAACCCCGTGGCGCGGACATCTACAAG 780
Db 721 CAGGAGCAGATCGCTGATGACCAAGCAACCCCGTGGCGCGGACATCTACAAG 780
QY 781 CGGTGATCATCTGGGCTTGAACAGATCGTGGGATGTACAGCCCGTGAGCATCTG 840
Db 781 CGGTGATCATCTGGGCTTGAACAGATCGTGGGATGTACAGCCCGTGAGCATCTG 840
QY 841 GACATCCCGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTTCTCAAGAC 900
Db 841 GACATCCCGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTTCTCAAGAC 900
QY 901 CTGCGCGCGAGAGAGCCACCAAGAGCTGAAGAGCTGATGACCAAGACCTGCTGTG 960
Db 901 CTGCGCGCGAGAGAGCCACCAAGAGCTGAAGAGCTGATGACCAAGACCTGCTGTG 960
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCGCGCGCGCACCTG 1020
Db 961 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCGCGCGCGCACCTG 1020
QY 1021 GAGAGATGATGACCGCTGCGAGCGCTGGGCGCGCGCGCGCACAGGCGCGCTGCTG 1080
Db 1021 GAGAGATGATGACCGCTGCGAGCGCTGGGCGCGCGCGCGCACAGGCGCGCTGCTG 1080
QY 1081 GCCGAGCGGATGAGCGGCAAGCGCTGAACATCATGATGACAGAGAGCACTTCAAG 1140
Db 1081 GCCGAGCGGATGAGCGGCAAGCGCTGAACATCATGATGACAGAGAGCACTTCAAG 1140
QY 1141 GCGCGCGCGGCAAGCTCAAGTCTTCACTGCGGCAAGAGGCGCACATCGCCCAAGAC 1200
Db 1141 GCGCGCGCGGCAAGCTCAAGTCTTCACTGCGGCAAGAGGCGCACATCGCCCAAGAC 1200
QY 1201 TGGCGCGCGCGCGCAAGAGGCTGTGAGAGTGGCGCAAGAGGCGCACAGATGAAG 1260
Db 1201 TGGCGCGCGCGCGCAAGAGGCTGTGAGAGTGGCGCAAGAGGCGCACAGATGAAG 1260
QY 1261 GACTGCAACGAGCGCAAGCTTCTGTGGGCAAGATCTGGGCCAGCCCAAGAGGCGCG 1320
Db 1261 GACTGCAACGAGCGCAAGCTTCTGTGGGCAAGATCTGGGCCAGCCCAAGAGGCGCG 1320
QY 1321 CCGGCAACTTCTGTGAGAACCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 1321 CCGGCAACTTCTGTGAGAACCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
```

```
QY 1381 CCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCG 1440
Db 1381 CCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCG 1440
QY 1441 CGGAGCCCTACCGCGAGAGCCCTGACCGCCCTGCGCAGCGCTGTGCGAGCGCGCGCTG 1500
Db 1441 CGGAGCCCTACCGCGAGAGCCCTGACCGCCCTGCGCAGCGCTGTGCGAGCGCGCGCTG 1500
QY 1501 AGCCAGTAA 1509
Db 1501 AGCCAGTAA 1509

RESULT 4
US-09-899-575-21
; Sequence 21, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
; APPLICANT: Zur Megele, Jan
; APPLICANT: Barnett, Susan W.
; APPLICANT: Egnelbrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janse
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP01631.102
; CURRENT APPLICATION NUMBER: US/09/899, 575
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475, 704
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag
; OTHER INFORMATION: coding sequence of HIV strain AF110967
US-09-899-575-21

Query Match 98.7%; Score 1489.8; DB 10; Length 1509;
Best Local Similarity 99.2%; Pred. No. 3.1e-313;
Matches 1497; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGAGAAAGTGAAGTGAAGATCCGC 60
Db 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGAGAAAGTGAAGTGAAGATCCGC 60
QY 61 CTGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACCTGTGTGGCGAGCGCGAG 120
Db 61 CTGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACCTGTGTGGCGAGCGCGAG 120
QY 121 CTGAGGAGCTTCCGCTGAACCCCGCGCTGTGAGAGCGCGAGGCTGCAAGCAGATC 180
Db 121 CTGAGGAGCTTCCGCTGAACCCCGCGCTGTGAGAGCGCGAGGCTGCAAGCAGATC 180
QY 181 ATGAAGCAGCTGACCGCGCTGACAGCGCGCACCGAGAGCTGCGACGCTGTACAAC 240
Db 181 ATGAAGCAGCTGACCGCGCTGACAGCGCGCACCGAGAGCTGCGACGCTGTACAAC 240
QY 241 ACCGTGACCACTTCTACTGCTGCAAGCGCGCATGAGGTCCGACACCAAGAGGCC 300
Db 241 ACCGTGACCACTTCTACTGCTGCAAGCGCGCATGAGGTCCGACACCAAGAGGCC 300
QY 301 CTGACAAAGATCGAGAGAGAGCAAGCAAGTCCAGCAGAGAGCCAGAGGCCAAGAG 360
Db 301 CTGACAAAGATCGAGAGAGAGCAAGCAAGTCCAGCAGAGAGCCAGAGGCCAAGAG 360
QY 361 GCCGAGCGAGGTGAGCGCAACTACCCCATCTGTGAGAACTGAGAGGCCAGATGTG 420
Db 361 GCCGAGCGAGGTGAGCGCAACTACCCCATCTGTGAGAACTGAGAGGCCAGATGTG 420
QY 421 CACGAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAGAGCC 480
```

```
Db 421 CACCAGGCCATCAGCCCCCGCACCCCTGGAACGCTGGGTGAAGTGAATCGAGAGAGGCC 480
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGGCGCCACCCCGCAGAC 540
Db 481 TTCAGCCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGGCGCCACCCCGCAGAC 540
QY 541 CTGAACACGATGTTGAACACCGTGGGGGCCACCAAGCCCGCATGAGATGCTGAAGAC 600
Db 541 CTGAACACCATGCTGAACACCGTGGGGGCCACCAAGCCCGCATGAGATGCTGAAGAC 600
QY 601 ACCATCAACGAGAGCGCCCGGAGTGGGACCGCTGCAACCCCGTGCAGCGCGCGCCGTG 660
Db 601 ACCATCAACGAGAGCGCCCGGAGTGGGACCGCTGCAACCCCGTGCAGCGCGCGCCGTG 660
QY 661 GCGCGCGCGCAGATGCGCGGACCCCGCGGCGAGCATCGCGCGCGCCACAGCACCTG 720
Db 661 GCGCGCGCGCAGATGCGCGGACCCCGCGGCGAGCATCGCGCGCGCCACAGCACCTG 720
QY 721 CAGGACGATCGCCTGATGACCAAGAACCCCGCTGCCGTGCCGCGACATCTAACAG 780
Db 721 CAGGACGATCGCCTGATGACCAAGAACCCCGCTGCCGTGCCGCGACATCTAACAG 780
QY 781 CGGTGATCATCTGGGCTTGAACAGATCGTGGGATGTACAGCCCGTGAGCATCTG 840
Db 781 CGGTGATCATCTGGGCTTGAACAGATCGTGGGATGTACAGCCCGTGAGCATCTG 840
QY 841 GACATCCGCGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAACCGCTTCTCAAGAC 900
Db 841 GACATCCGCGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAACCGCTTCTCAAGAC 900
QY 901 CTGGCGCGCGCAGAGCGCCACCAAGAGCTGAAGACTGGATGACCGGACCTGCTGTG 960
Db 901 CTGGCGCGCGCAGAGCGCCACCAAGAGCTGAAGACTGGATGACCGGACCTGCTGTG 960
QY 961 CAGAACGCCAACCCCGCACTGCAAGACCATCTGCGCGCTCTCGCGCGCGCGCGCACCTG 1020
Db 961 CAGAACGCCAACCCCGCACTGCAAGACCATCTGCGCGCTCTCGCGCGCGCGCGCACCTG 1020
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCGCGCGCACAGCGCGCGCTG 1080
Db 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCGCGCGCACAGCGCGCGCTG 1080
QY 1081 GCGGAGCGGATGAGCGCAAGGCCAACAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140
Db 1081 GCGGAGCGGATGAGCGCAAGGCCAACAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140
QY 1141 GCGCGCGCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGGCGCCATCGCCAAAGAC 1200
Db 1141 GCGCGCGCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGGCGCCATCGCCAAAGAC 1200
QY 1201 TGGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGGCAAGGAGGAGGAGGAGGAGGAGGAG 1260
Db 1201 TGGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGGCAAGGAGGAGGAGGAGGAGGAG 1260
QY 1261 GACTGCAACGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGGCCAGCCCAAGGGCGCG 1320
Db 1261 GACTGCAACGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGGCCAGCCCAAGGGCGCG 1320
QY 1321 CCGGCGCAACTTCTGCAAGAACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 1321 CCGGCGCAACTTCTGCAAGAACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
QY 1381 CCGGCGCGAGCTTCGCTTGAAGAGAACACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
Db 1381 CCGGCGCGAGCTTCGCTTGAAGAGAACACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
QY 1441 CCGGAGCCCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
Db 1441 CCGGAGCCCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
QY 1501 AGCCAGTAA 1509
Db 1501 AGCCAGTAA 1509
```

```
Db 1501 AGCCAGTAA 1509
RESULT 5
US-09-899-575-51
; Sequence 51, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
; APPLICANT: Zur Megele, Jan
; APPLICANT: Barnett, Susan W.
; APPLICANT: Egnelbrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janse
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP01631.102
; CURRENT APPLICATION NUMBER: US/09/899,575
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475,704
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HIV Type C
US-09-899-575-51
```

```
Query Match 86.0%; Score 1297.6; DB 10; Length 1494;
Best Local Similarity 92.7%; Pred. No. 1.3e-271;
Matches 1402; Conservative 0; Mismatches 89; Indels 21; Gaps 3;
```

```
QY 1 ATGGCGCGCGCGCGCAGCATCTCGCGCGCGCGCAGAGCTGACAAAGTGGAGAGATCCG 60
Db 1 ATGGCGCGCGCGCGCAGCATCTCGCGCGCGCGCAGAGCTGACAAAGTGGAGAGATCCG 60
QY 61 CTGGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCGAGCGCGAG 120
Db 61 CTGGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCGAGCGCGAG 120
QY 121 CTGAGGGCTTCGCTTGAACCCCGGCTGTGAGAACCGCGAGCGCGGCTGCAAGCATC 180
Db 121 CTGAGGGCTTCGCTTGAACCCCGGCTGTGAGAACCGCGAGCGCGGCTGCAAGCATC 180
QY 181 ATGAAGAGCTGACCGCGCGCGCGCGCAGACCGGAGCGGAGAGTGGCGAGCGCTGTAC 240
Db 181 ATGAAGAGCTGACCGCGCGCGCGCGCAGACCGGAGCGGAGAGTGGCGAGCGCTGTAC 240
QY 241 ACCGTGGCCACCTGTACTGCTGAGCGCGGCGATCGAGGTCCGCGACCAAGAGGCG 300
Db 241 ACCGTGGCCACCTGTACTGCTGAGCGCGGCGATCGAGGTCCGCGACCAAGAGGCG 300
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCAGCAGAGAAGACCAAGAGGCGAGAG 360
Db 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCAGCAGAGAAGACCAAGAGGCGAGAG 360
QY 361 GCCGACGCAAGGTGAGCCAGAACTACCCCATCTGCAAGAACTGAGGGCGAGAGGTG 420
Db 361 GCCGACGCAAGGTGAGCCAGAACTACCCCATCTGCAAGAACTGAGGGCGAGAGGTG 420
QY 421 CACCAAGGCATCAGCCCCCGCAGACCTTGAACGCTGGGTGAAGTGAATCGAGAGAGCC 480
Db 421 CACCAAGGCATCAGCCCCCGCAGACCTTGAACGCTGGGTGAAGTGAATCGAGAGAGCC 480
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCAACCGCTTGAAGAGGCGCGCGCGCGCGCGAG 540
Db 481 TTCAGCCCCGAGGTGATCCCATGTTCAACCGCTTGAAGAGGCGCGCGCGCGCGCGAG 540
QY 541 CTGAACACGATGTTGAACACCGTGGGGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCG 600
Db 541 CTGAACACCATGCTGAACACCGTGGGGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCG 600
```

QY	601	ACCATCAACGAGAGGCGCCGAGTGGGACCGCCTGACCCCGTGAAGCCCGGCGCCGCTG	660
Db	601	ACCATCAACGAGAGGCGCCGAGTGGGACCGCACCCACCCCGTGAAGCCCGGCGCCGCTG	660
QY	661	GCCCCGGCCAGATGCGCGGACCCCGCGGCGAGCGACATCGCCGGCGCACCGACCCCTG	720
Db	661	GCCCCGGCCAGATGCGCGGAGCCCGCGGCGAGCGACATCGCCGGCACCGACCGCCCTG	720
QY	721	CAGAGCAGATCGCCTGGATGACCAGAACCCCCCGTGCCCGTGGCGACATTTACAAG	780
Db	721	CAGAGCAGATCGCCTGGATGACCAGAACCCCCCGTATCCCGTGGAGGACATTTACAAG	780
QY	781	CGGTGATCATCTTGCGGCTGAAACAAGATCGTGCGGATGTACAGCCCGTGAGCATCTG	840
Db	781	CGGTGATCATCTTGCGGCTGAAACAAGATCGTGCGGATGTACAGCCCGTGAGCATCTG	840
QY	841	GACATCCGCCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTTCAAGACC	900
Db	841	GACATCAAGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTTCAAGACC	900
QY	901	CTGCGCGCCGAGCAGGCCACCCAGGACGTGAAGAACTGGATGACCGGACCCCTGTGTTG	960
Db	901	CTGCGCGCCGAGCAGGCCACCCAGGACGTGAAGAACTGGATGACCGGACACCCTGTGTTG	960
QY	961	CAGAACGCCAACCCCGGACTGCAAGACCATCTGCGCGCTCTCGGCCCCGGCGCCACCCTG	1020
Db	961	CAGAACGCCAACCCCGGACTGCAAGACCATCTGCGCGCTCTCGGCCCCGGCGCCACCCTG	1020
QY	1021	GAGGAGTATGACCGCCTGCCAGGGCGTGGCGGCCCCGGCCACAAGGCCCGCTGCTG	1080
Db	1021	GAGGAGTATGATGACCGCCTGCCAGGGCGTGGCGGCCCCGACCAAGGCCCGCGTGTG	1080
QY	1081	GCCGAGGCGATGAGGCCAGGCCAACGCTGAACATCATGATGCAAGAGCAACTTCAAG	1140
Db	1081	GCCGAGGCGATGAGGCCAGGCCAACG--AACTCTGTGTGACGCGCAACTTCAAG	1137
QY	1141	GCGCCCCGCGCAACGTCAAGTGTCTCACTGCGGCAAGAGGCCACATCGCCAAAGAC	1200
Db	1138	GCGCAACAACCGCATCATCAAGTGTCTCACTGCGGCAAGTGGCCACATCGCCGCAAC	1197
QY	1201	TGCGCGCGCCCCCGCAAGAGGCGTGTGAAGTGGCGGCAAGAGGGGCCACAGATGAAG	1260
Db	1198	TGCGCGCGCCCCCGCAAGAGGCGTGTGAAGTGGCGGCGGCAAGAGGGGCCACAGATGAAG	1257
QY	1261	GACTGCACCGAGCGCCAGGCAACTTCTTGCGCAAGATTTGGGCCAGCCACAAGGGCGCG	1320
Db	1258	GACTGCACCGAGCGCCAGGCAACTTCTTGCGCAAGATTTGGGCCAGCCACAAGGGCGCG	1317
QY	1321	CCCGCAACTTCTTGCAAGACCGCAGCAGC---CGCGCGCCCCACCGTGCCACCGCGC	1377
Db	1318	CCCGCAACTTCTTGCAAGACCGCGCCGAGCCACCGCCCCCGCGAGCCACCGCGC	1377
QY	1378	CCCCCGCGCGAGCTTCCGCTTGAAGAGACACCCCGCGCCCCCAAGCAGAGCCCAAG	1437
Db	1378	CCCCCGCGCGAGCTTCCGCTTGAAGAGACACCCCGTGCCCCCAAGAGAGAGAG	1437
QY	1438	GACCGGAGCCCTACCGCGAGCGCCCTGACCGCCCTGCGCAGCCTGTTGGCAGCGGCCCC	1497
Db	1438	-----CGCGAGCCCTGACCGCCTGAAGAGCCTGTTGGCAGCGACCCC	1482
QY	1498	CTGAGCCAGTAA	1509
Db	1483	CTGAGCCAGTAA	1494

```

; APPLICANT: Egnelbrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janse
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP01631.102
; CURRENT APPLICATION NUMBER: US/09/899,575
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475,704
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 99
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Gag_TV2_C_Zaoptb
US-09-899-575-99

```

Query Match	85.8%;	Score 1294.8;	DB 10;	Length 1491;
Best Local Similarity	92.7%;	Pred. No. 5.2e-271;		
Matches 1401; Conservative	0;	Mismatches 87;	Indels 24;	Gaps 3;

QY	1	ATGGGCGCCGCGCCAGCATCTGCGCGCGGAGAAGCTGGA CAAGTGGAGAAAGATCCGC	60
Db	1	ATGGGCGCCGCGCCAGCATCTGCGCGCGCGCAAGCTGGA CAAGTGGAGAAAGATCCGC	60
QY	61	CTGCGCCCGCGCGCAAGAAAGCACTAATGCTGAAGCACTGTGTGGGCCAAGCCGCGAG	120
Db	61	CTGCGCCCGCGCGCGCAAGCACTAATGCTGAAGCACTGTGTGTGGGCCAAGCCGCGAG	120
QY	121	CTGAGGAGCTTGCCTGTGAACCCCGGCTGTGTGAGACCGCGAGGAGCTGTGAAGCAGATC	180
Db	121	CTGAGGAGCTTGCCTGTGAACCCCGGCTGTGTGAGACCAAGCAGGCTGTGCGCAGATC	180
QY	181	ATGAAGCAGCTGCAGCCCGCCTGTCAAGCCGGCACCGAAGAGCTGCGCAGCTGTACAAC	240
Db	181	ATCAAGCAGCTGCAGCCCGCCTGTCAAGCCGGCACCGAAGAGATCCGCAAGCTGTTCAC	240
QY	241	ACCGTGGCCACCCCTGTACTGCGTGCAGCGCCGGCATCGAGTCCGCGACACCAAGAGGCC	300
Db	241	ACCGTGGCCACCCCTGTACTGCGTGCAGAGGGCATCGACGTGCGGACACCAAGAGGCC	300
QY	301	CTGCAAGAATCGAGGAGGAGCAGAA CAAGTCCCAAGCAAGAACCCAGAGGCCAAGAGAG	360
Db	301	CTGCAAGAATCGAGGAGGAGCAGAA CAAGTCCCAAGCAAGAACCCAGAGGCCAAGAGGCC	360
QY	361	GCCGACGGCAAGGTGAGCCAGAACTAACCCCATCGTGCAGAACCTGCAGGGCCAGATGTTG	420
Db	361	GCCGACAAAGAGGTGAGCCAGAACTAACCCCATCGTGCAGAACCTGCAGGGCCAGATGTTG	420
QY	421	CACCAAGGCATCAGCCCCCGCACCTGTGAACGCTGGGTGAAGGTGATCGAGGAAAGGCC	480
Db	421	CACCAAGGCATCAGCCCCCGCACCTGTGAACGCTGGGTGAAGGTGATCGAGGAAAGGCC	480
QY	481	TTCAGCCCCGAGGTGATCCCCCATGTTCAACCGCCCTGAAGGAGGCGCCACCCCCAGGAC	540
Db	481	TTCAGCCCCGAGGTGATCCCCCATGTTCAACCGCCCTGAAGGAGGCGCCACCCCCAGGAC	540
QY	541	CTGAACAACGATGTTGAACACCGTGGCGCGCCACCAAGGCCGCATGCAATGCTGAAGGAC	600
Db	541	CTGAACAACGATGTTGAACACCGTGGCGCGCCACCAAGGCCGCATGCAATGCTGAAGGAC	600
QY	601	ACCATCAACGAGGAGCGCCCGGAGTGGGACCGGCTGCACCCCGTGAAGGCCCGGCCGTG	660
Db	601	ACCATCAACGAGGAGCGCCCGGAGTGGGACCGGCTGCACCCCGTGAAGGCCCGGCCGTG	660
QY	661	GCCCCCGGCAGATGCGCGACCCCGCGGAGCGACATGCGCGGCACCAAGCAACCTTG	720
Db	661	GCCCCCGGCAGATGCGCGAGCCCGCGGAGCGACATGCGCGGCACCAAGCAACCTTG	720
QY	721	CAGGAGCAGATCGCCTGATGACAGCAACCCCCCGTGCCTGTGGCGACATCTACAAG	780
Db	721	CAGGAGCAGATCGCCTGATGACAGCAACCCCCCGTGCCTGTGGCGACATCTACAAG	780

Db	721	CAGGAGCAGATCGCCTGAGATGACCAAGCAACCCCCCATCTCCCGTGGCGCATCTACAAG	780
QY	781	CGGTGATCATCTTGGGCGCTGAAACAAGATCGTGGCATGTACAGCCCCGTGAGCATCTTG	840
Db	781	CGCTGATCATCTTGGGCGCTGAAACAAGATCGTGGCATGTACAGCCCCGTGAGCATCTTG	840
QY	841	GACATCCGCGAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGACC	900
Db	841	GACATCAAGCAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGACC	900
QY	901	CTGCGCGCGGAGCAGGGCCACCCAGAGAGCTGTAGAACTGGATGACCGAGACCCTGCTGTTG	960
Db	901	CTGCGCGCGGAGCAGAGCAACCAGAGAGTGTAGAACTGGATGACCGAGACCCTGCTGTTG	960
QY	961	CAGAACGCCAACCCCGACTGTCAAGACCATCTGCGGCTCTCGGCCCCGGGCGCCACCCTTG	1020
Db	961	CAGAACGCCAACCCCGACTGTCAAGACCATCTGCGGCTCTCGGCCCCGGGCGCCAGCCTTG	1020
QY	1021	GAGGAGATGATGACCCGCTGCGAGGGCGGTGGGCGGCTCCCGGCCAACAGCCCCGCGTCTG	1080
Db	1021	GAGGAGATGATGACCCGCTGCGAGGGCGGTGGGCGGCTCCAGGCCAACAGCCCCGCGTCTG	1080
QY	1081	GCCGAGGCGATGAGCCAGGCCCAACAGCGTGAACATCATGATGACAGAAAGCAACTTCAAAG	1140
Db	1081	GCCGAGGCGCATGAGCCAGGCCCAACACACAGCGTGTATGATCCAGAAAGCAACTTCAAAG	1140
QY	1141	GGCCCCCGGCGCAACGTCAAGTGTCTCAACTGCGGCAAGGAGGGCCATCGCCAAAGAAC	1200
Db	1141	GGCCCCCGGCGCGCGCGTGAAGTGTCTCAACTGCGGCGGAGAGGGCCATCGCCGCAAC	1200
QY	1201	TGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGGCAAGGAGGGCCACAGATGAAG	1260
Db	1201	TGCGCGCGCCCCCGCAAGCGCGGTGTGGAAGTGGGCAAGGAGGGCCACAGATGAAG	1260
QY	1261	GACTGCACGAGCGGCCAGGCCCACTTCTTGGGCAAGATCTGGCCCAAGGCCCGC	1320
Db	1261	GACTGCACGAGCGGCCAGGCCCACTTCTTGGGCAAGATCTGGCCCAAGGGCGCG	1320
QY	1321	CCCGGCACTTCTGTGCAGAACCGCAGCGAGCCCGCGCCCCCAACCC--GTGCCACCGCG	1377
Db	1321	CCCGGCACTTCTGTGCAAGCGCGCCCGAGCCCAACCGCCCCCTGGAAGCCACCGCG	1380
QY	1378	CCCCCGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGCCCCCAAGCAGAGGCCCAAG	1437
Db	1381	CCCCCGCGGAGAGCTTCAAGTTCAAGAGA-----CCCCCAAGCAGAGGCCCAAG	1431
QY	1438	GACCGCGAGCCCTACCGCGAGCGCTGACCGCGCTGCGCAGCCTGTTCGGCAGCGGCC	1497
Db	1432	G-----ACCGCGAGCGCTGACCAAGCCTGAAGAGCCTGTTCGGCAGCGGCC	1479
QY	1498	CTGAGCCAGTAA 1509	
Db	1480	CTGAGCCAGTAA 1491	

```

RESULT 7
US-09-967-464-63
; Sequence 63, Application US/09967464
; Publication No. US20030138453A1
; GENERAL INFORMATION:
; APPLICANT: O'Hagan, Derek
; APPLICANT: Otten, Gillis
; APPLICANT: Donnelly, John J.
; APPLICANT: Polo, John M.
; APPLICANT: Barnett, Susan
; APPLICANT: Singh, Mamohan
; APPLICANT: Ulmer, Jeffrey
; APPLICANT: Dubensky, Jr., Thomas W.
; TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF HETEROLOGOUS NUCLEIC ACIDS
; FILE REFERENCE: P16269.004
; CURRENT APPLICATION NUMBER: US/09/967,464
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/236,105

```

```

! PRIOR FILING DATE: 2000-09-28
! PRIOR APPLICATION NUMBER: 60/315,905
! PRIOR FILING DATE: 2001-08-30
! NUMBER OF SEQ ID NOS: 68
! SOFTWARE: PatentIn version 3.1
! SEQ ID NO 63
! LENGTH: 1479
! TYPE: DNA
! ORGANISM: Human immunodeficiency virus type 1
US-09-967-464-63

```

Query Match	84.6%;	Score 1276.8;	DB 10;	Length 1479;
Best Local Similarity	92.5%;	Pred. No. 4.1e-267;		
Matches 1399; Conservative	0;	Mismatches 77;	Indels 36;	Gaps 4;

QY	1	ATGGGCGCCCGCCAGCATCCTCGCGCGCGGAGAACTGGACAAGTGGAGAAGATCCGC	60
Db	1	ATGGGCGCCCGCCAGCATCCTCGCGCGCGGAGAACTGGACAAGTGGAGAAGATCCGC	60
QY	61	CTGCGCCCCCGCGCAAGAAAGCACTACATGCTGAAGCACTGTGTGTGGCCAGCCCGAG	120
Db	61	CTGCGCCCCCGCGCAAGAAAGTGTACATGATGAAGCACTGTGTGTGGCCAGCCCGAG	120
QY	121	CTGGAAGGGCTTCGCGCTTGAAACCCCGGCTGTGTGAGACCGCCGAGGCTGCAGCAGATC	180
Db	121	CTGGAAGAAGTTCGCGCTTGAAACCCCGGCTGTGTGAGACCAAGGAGGCTGCAGCAGATC	180
QY	181	ATGAAGCAGCTGCGACCCCGCTGCAGACCCGGCACCGAGGAGCTGCGCAGCTGTACAC	240
Db	181	ATCCGCCAGCTGCAACCCCGCTGCAGACCGGCAGCGAGGAGCTGAAGAGCTGTTCAC	240
QY	241	ACCGTGGCCACCTCTGTACTGCGGTGACGCGCGCATCGAGGTCCGCGCACCAAGAGAGCC	300
Db	241	ACCGTGGCCACCTGTACTGCGGTGACGAGAAAGATCGAGGTCCGCGCACCAAGAGAGCC	300
QY	301	CTGGAACAAGATCGAGAGGAGCAGAAACAAGTCCCAAGACAAGACCAAGAGCCCAAGAG	360
Db	301	CTGGAACAAGATCGAGAGGAGCAGAAACAAGTCCAGACAGAAGATCCAGAGGCCAGAGCC	360
QY	361	GCCGAC--GGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACCTGCAGGGCCAGATG	417
Db	361	GCCGACAAAGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGCAGGGCCAGATG	420
QY	418	GTCACACGAGCCATCAGCCCCCGCACCTGAACGCGCTGGTGAAAGGTGATCGAGAGAG	477
Db	421	GTCACACGAGCCATCAGCCCCCGCACCTGAACGCGCTGGTGAAAGGTGATCGAGAGAG	480
QY	478	GCGTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCTGAGCGAGGCGCCACCCCGAG	537
Db	481	GCGTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCTGAGCGAGGCGCCACCCCGAG	540
QY	538	GACCTGAACACGATGTTGAACACCGTGGGCGGCCAACAGGCCGCAATGCAGATGCTGAAG	597
Db	541	GACCTGAACACGATGTTGAACACCGTGGGCGGCCAACAGGCCGCAATGCAGATGCTGAAG	600
QY	598	GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGCTGCAGGCGCGGCC	657
Db	601	GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGCTGCAGCGCGGCC	660
QY	658	GTCGCCCCCGCGCAGATGCGCGACCCCGCGCGCAGCATCGCCGCGCCACCAAGCAAC	717
Db	661	ATCGCCCCCGCGCAGATGCGCGACCCCGCGCGCAGCATCGCCGCGCCACCAAGCAAC	720
QY	718	CTGCAAGAGCAGATCGCGCTGATGACCAAGCAACCCCGCTGCGCGTGGGCGCATCTAC	777
Db	721	CTGCAAGAGCAGATCGCGCTGATGACCAAGCAACCCCGCATCCCGTGGGCGCATCTAC	780
QY	778	AAGCGGTGATCATCTGGGCTGAACAAGATCGTGGCGGATGTACAGCCCGGTGAGATC	837
Db	781	AAGCGGTGATCATCTGGGCTGAACAAGATCGTGGCGGATGTACAGCCCGGTGAGATC	840
QY	838	CTGAACATCCGCAAGGCGCCCAAGAGGCCCTTCCGCGACTACGTGACCGGCTTCTTCAAG	897

Db 841 CTGACATCAAGAGGGCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAG 900  
QY 898 ACCCTGCGCGCGAGAGGCCACCCAGAGCGTGAAGACTGATGACGAGACCTGCTG 957  
Db 901 ACCCTGCGCGCGAGAGGCCACCCAGAGGTGAAGACTGATGACGAGACCTGCTG 960  
QY 958 GTGCAGAACGCCCAACCCGCACTGCAAGACCATCTGCGCGCTCTGCGCCCGCGCAAC 1017  
Db 961 GTGCAGAACGCCCAACCCGCACTGCAAGACCATCTGCGCGCTCTGCGCCCGCGCAAC 1020  
QY 1018 CTGAGAGAGATGATGACCGCTGCGAGGGCGTGGCGCGCCCGCCCAAGGCCCGCTG 1077  
Db 1021 CTGAGAGAGATGATGACCGCTGCGAGGGCGTGGCGCGCCCGCCCAAGGCCCGCTG 1080  
QY 1078 CTGCGCGAGCGGATGAGCCAGGCCAACAGCGTGAACATCATGATGAGAGAGCACTTC 1137  
Db 1081 CTGCGCGAGCGGATGAGCCAGGCCAACAC--AGCGTATGATGAGAGAGCACTTC 1137  
QY 1138 AAGGGCCCCCGCGCAAGTCAAGTCTTCACTGCGGCAAGAGAGGCCACATCGCCAAG 1197  
Db 1138 AAGGGCCCCCGCGCAAGTCTTCAAGTCTTCACTGCGGCAAGAGAGGCCACATCGCCGC 1197  
QY 1198 AACTGCGCGCGCGCGCAAGAGGGCTGCTGMAAGTGGCGGCAAGAGGGCCACAGATG 1257  
Db 1198 AACTGCGCGCGCGCGCAAGAGGGCTGCTGMAAGTGGCGGCAAGAGGGCCACAGATG 1257  
QY 1258 AAGGACTGACCGAGCGCCAGGCCAATTCTGCGCAAGATCTGGCCCAAGCCCAAGGGC 1317  
Db 1258 AAGGACTGACCGAGCGCCAGGCCAATTCTGCGCAAGATCTGGCCCAAGCCCAAGGGC 1317  
QY 1318 CGCCCCGCGCAACTTCTGAGAACCGGAGCGCGCGCCCGCCCAAGCTGCGCCACCGCC 1377  
Db 1318 CGCCCCGCGCAACTTCTGAGAGCG--CGCCGAGCCCAAGCTGCGCCACCGCC 1359  
QY 1378 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCCCAAGAGAGGCCAAG 1437  
Db 1360 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCCCAAGAGAGGCCAAG 1419  
QY 1438 GACCGCGAGCGCTTACCGCGAGCGCCCTGACCGCGCGAGCGCTGTTGCGAGCGCCCC 1497  
Db 1420 G-----ACCGGAGAGCCCTGACCGAGCTGAAGAGCTGTTGCGCAAGCAACCC 1467  
QY 1498 CTGAGCCAGTAA 1509  
Db 1468 CTGAGCCAGTAA 1479

RESULT 8  
US-09-899-575-3

/ Sequence 3, Application US/09899575  
/ Publication No. US20030223961A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Zur Megede, Jan  
/ APPLICANT: Barnett, Susan W.  
/ APPLICANT: Egnelbrecht, Susan  
/ APPLICANT: van Rensburg, Estrelita Janse  
/ TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
/ TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
/ FILE REFERENCE: PP01631.102  
/ CURRENT APPLICATION NUMBER: US/09/899,575  
/ CURRENT FILING DATE: 2001-07-05  
/ PRIOR APPLICATION NUMBER: 09/475,704  
/ PRIOR FILING DATE: 1999-12-30  
/ NUMBER OF SEQ ID NOS: 135  
/ SOFTWARE: Patent Ver. 2.0  
/ SEQ ID NO 3  
/ LENGTH: 1479  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag  
/ OTHER INFORMATION: of HIV strain AF110965  
US-09-899-575-3

Query Match 84.6%; Score 1276.8; DB 10; Length 1479;  
Best Local Similarity 92.5%; Pred. No. 4.1e-267;  
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;  
QY 1 ATGGGCGCGCGCGCGAGCATCTCTGCGCGCGGAGAACTGGACAAAGTGGAGAAAGATCCGC 60  
Db 1 ATGGGCGCGCGCGCGAGCATCTCTGCGCGCGGAGAACTGGACAAAGTGGAGAAAGATCCGC 60  
QY 61 CTGCGCGCGCGCGCGCAAGAGGCTACATGATGAAGCACTTGTGTGGCCAGCGCGAG 120  
Db 61 CTGCGCGCGCGCGCGCAAGAGGCTACATGATGAAGCACTTGTGTGGCCAGCGCGAG 120  
QY 121 CTGAGAGGCTTCCGCTTGAACCCCGCGCTGCTGGAAGACCGCGGAGGGCTGCAAGCATC 180  
Db 121 CTGAGAGAGTTCCGCTTGAACCCCGCGCTGCTGGAAGACCGCGGAGGGCTGCAAGCATC 180  
QY 181 ATGAAGCAGCTGCAAGCCCGCTGCAAGCCGCGACCGAGAGGCTGCGAGCTGTACAAC 240  
Db 181 ATCCGCGAGCTGCAAGCCCGCTGCAAGCCGCGACCGAGAGGCTGGAAGGCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCTGCTGCAAGCCGCGATCGAGGTCCGCGACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTACTGCTGCTGCAAGAGAGATCGAGGTCCGCGACCAAGAGGCC 300  
QY 301 CTGGAACAAGATGAGAGAGAGGAGCAAGTCCAGCAGAGAACCCAGAGGCCAAGAG 360  
Db 301 CTGGAACAAGATGAGAGAGAGGAGCAAGTCCAGCAGAGATCCAGAGGCCAAGAGGCC 360  
QY 361 GCGGAC--GGCAAGGTGAGCCGAGAACTACCCCATCTGTCAGAACTGACAGGCGCAGATG 417  
Db 361 GCGGACAAAGGAGAGGTGAGCCAGAACTACCCCATCTGTCAGAACTGACAGGCGCAGATG 420  
QY 418 GTGCAACAGGCGCATCAGCCCGCGCAACCTGTAACGCTGCTGTAAGGTATGAGAGAGAG 477  
Db 421 GTGCAACAGGCGCATCAGCCCGCGCAACCTGTAACGCTGCTGTAAGGTATGAGAGAGAG 480  
QY 478 GCCTTACGCGCGGAGGTGATCCCATGTTTACCGCGCGCTGAGCGAGGGCGCCACCCCGAG 537  
Db 481 GCCTTACGCGCGGAGGTGATCCCATGTTTACCGCGCGCTGAGCGAGGGCGCCACCCCGAG 540  
QY 538 GACCTGAACAGATGTTGAACACCGTGGCGGCGCAACAGGCGCCATGACATGCTGAAG 597  
Db 541 GACCTGAACAGATGTTGAACACCGTGGCGGCGCAACAGGCGCCATGACATGCTGAAG 600  
QY 598 GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCAACCCCGTGAGCGCGCGCC 657  
Db 601 GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCAACCCCGTGAGCGCGCGCC 660  
QY 658 GTGGCGCGCGCGCGAGTGGCGGACCCCGCGCGGAGCGACATCGCGCGCGCACCAAGCACC 717  
Db 661 ATCGCGCGCGCGCGAGTGGCGGAGCCCGCGCGGAGCGACATCGCGCGCGCACCAAGCACC 720  
QY 718 CTGCAAGAGAGATGCGCTGATGACCAAGACCCCGCTGCGCGACTACGTGAGCCGCTTCAAG 777  
Db 721 CTGCAAGAGAGATGCGCTGATGACCAAGACCCCGCTGCGCGACTACGTGAGCCGCTTCAAG 780  
QY 778 AAGCGGTGATCATCTGGGCTGTAACAAAGATGTCGCGATGTACAGCCCGTGAGCATC 837  
Db 781 AAGCGGTGATCATCTGGGCTGTAACAAAGATGTCGCGATGTACAGCCCGTGAGCATC 840  
QY 838 CTGACATCCGCGAGGCGCGCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTCAAG 897  
Db 841 CTGACATCAAGCAGGCGCGCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTCAAG 900  
QY 898 ACCCTGCGCGCGAGCAGGCCACCCAGAGCGTGAAGAACTGATGACCGACCTGCTG 957  
Db 901 ACCCTGCGCGCGAGCAGGCCACCCAGAGCGTGAAGAACTGATGACCGACCTGCTG 960  
QY 958 GTGCAGAACGCCAACCAGACTGCAAGACATCTGCGGCTCTGCGCGCGCGCGCAGC 1017  
Db 961 GTGCAGAACGCCAACCAGACTGCAAGACATCTGCGGCTCTGCGCGCGCGCGCAGC 1020

QY	1018	CTGGAGGAGATGATGACCGCGCTTGCCAGAGGGCGGTGGGCGGCCCCCGGCACCAAGGCCGCGTG	1077
Db	1021	CTGGAGGAGATGATGACCGCGCTTGCCAGAGGGCGGTGGGCGGCCCCCGGCACCAAGGCCGCGTG	1080
QY	1078	CTGGCCGAGGGCGATGAGCCAGGCGCAACAGCCGTGAATCATGTATGCAAGAAAGCAACTTC	1137
Db	1081	CTGGCCGAGGGCGATGAGCCAGGCGCAACACC--AGCGTATGATGTCAGAAAGCAACTTC	1137
QY	1138	AAGGGCCCCCGGCGCAACGTCAAGTGTCTCAACTGCGGCAAGGAGGGCCACATCGCCAAAG	1197
Db	1138	AAGGGCCCCCGGCGCATCGTCAAGTGTCTCAACTGCGGCAAGGAGGGCCACATCGCCCCG	1197
QY	1198	AACTGCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCAGGCAAGAGGGCCACCAAGATG	1257
Db	1198	AACTGCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCAGGCAAGAGGGCCACCAAGATG	1257
QY	1258	AAGGACTGCACCGAGCGCCAGGCCCACTTCTCTGGGCAAGATCTGGCCCAAGCCCAAGAGGC	1317
Db	1258	AAGGACTGCACCGAGCGCCAGGCCCACTTCTCTGGGCAAGATCTGGCCCAAGCCCAAGAGGC	1317
QY	1318	CGCCCCGGCAACTTCTCTGCAAGACCGAGCGAGCCGCGCCCCCACCGTGCCCAACCGCC	1377
Db	1318	CGCCCCGGCAACTTCTCTGCAAGCGG-----CCCCGAGGCCACCGCC	1359
QY	1378	CCCCCGCGCAGAGCTTCGCTTCGAGGAGACCACCCCCGCCCCCAAGCAGAGGCCCAAG	1437
Db	1360	CCCCCGCGCAGAGCTTCGCTTCGAGGAGACCACCCCCGGCCAAGAAAGCAGAGAGCAAG	1419
QY	1438	GACCGCGAGCCCTAACCGCGAGGCCCTGACCGCCCTGCGAGCCCTGTTGCGAGCGGCCCC	1497
Db	1420	G-----ACCGGAGACCTTGACCAAGCCTGAAGAGCCTGTTCGCGCAAGACCC	1467
QY	1498	CTGAGCCAGTAA	1509
Db	1468	CTGAGCCAGTAA	1479

```

RESULT 9
US-10-190-435-18
; Sequence 18, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18
; LENGTH: 3162
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagRTmut_C
US-10-190-435-18

```

Query Match	84.5%;	Score 1275.2;	DB 15;	Length 3162;
Best Local Similarity	92.5%;	Pred. No. 8.9e-267;		
Matches 1398;	Conservative	0;	Mismatches 78;	Indels 36; Gaps 4;

  

QY	1	ATGGGCGCCCGCGCCAGCATCTCTGCGCGCGGAGAGCTGACACAAGTGGAGAGAAATCCGC	60
DB	7	ATGGGCGCCCGCGCCAGCATCTCTGCGCGCGGAGAGCTGACAGCCCTGGAGAGCATCCGC	66
QY	61	CTGCGCCCCCGCGCGCAAGAGCACTACATGCTGAAGCACCTGGTGTGGGCCAGCCGGCAG	120
DB	67	CTGCGCCCCCGCGCGCAAGAGTGTACATGATGAAGCACCTGGTGTGGGCCAGCCGGCAG	126

OY	121	CTGAGAGGCTTCGCTCCCTGAAACCCCGGCTGCTGAGACCGCGGAGGGCTGCAAGCAGATC	180
Db	127	CTGAGAGAAATTCCGCTTGAACCCCGGCTGCTGAGAACCAAGGAGGGCTGCAAGCAGATC	186
OY	181	ATGAAGCAGCTGCAGACCCCGCTGCAGACCGGCAACGAGAGCTGCGCAGCTGTACAAC	240
Db	187	ATCCGCCAGCTGCACCCCGCTGCAGACCGGCAAGGAGAGCTGAAGAGCTGTTC AAC	246
OY	241	ACCGTGGCCACCTGTACTGCTGTCACGCGGCGATCGAGGTCGCGACACCAAGAGGCC	300
Db	247	ACCGTGGCCACCTGTACTGCTGTCACGAGAGATGAGGTCGCGACACCAAGAGGCC	306
OY	301	CTGGAACAAGATCGAGGAGGAGCAAACTGTCAGCAGAGAACCAGCAGGCCCAAGAG	360
Db	307	CTGGAACAAGATCGAGGAGGAGCAAACTGTCAGCAGAGAATCCAGCAGGCCGAGGCC	366
OY	361	GCCGAC--GGCAAGGTAGGCGAAGACTACCCCATCTGTGCAAACTGCAAGGCCAGATG	417
Db	367	GCCGACAAGGGCAAGGTAGGCGAAGACTACCCCATCTGTGCAAACTGCAAGGCCAGATG	426
OY	418	GTCGACCAAGGCGCATCAGCCCCCGCACCCCTGAAACGCTGGGTGAAGGTGATGAGAGAAG	477
Db	427	GTCGACCAAGGCGCATCAGCCCCCGCACCCCTGAAACGCTGGGTGAAGGTGATGAGAGAAG	486
OY	478	GCGTTCAAGCCCCGAGGTGATCCCCATGTTCAACGCGCTGAGCGAGGGCGGCCACCCCCAG	537
Db	487	GCGTTCAAGCCCCGAGGTGATCCCCATGTTCAACGCGCTGAGCGAGGGCGGCCACCCCCAG	546
OY	538	GACCTGAACACGATGTTGAACACCGTGGCGGACCACAGGCGGCCCATGCAATGCTGAAG	597
Db	547	GACCTGAACACGATGTTGAACACCGTGGCGGACCACAGGCGGCCCATGCAATGCTGAAG	606
OY	598	GACACCATCAACGAGGAGGCGCGGAGTGGGACCGGCTGCACCCCGTGCAGCGCGGCCCC	657
Db	607	GACACCATCAACGAGGAGGCGCGGAGTGGGACCGGCTGCACCCCGTGCAGCGCGGCCCC	666
OY	658	GTCGCCCCCGGCGCAGATGCGCGCAACCCCGCGGCAAGCAATCGCGCGGCCACAGCAC	717
Db	667	ATCGCCCCCGGCGCAGATGCGCGCAACCCCGCGGCAAGCAATCGCGCGGCCACAGCAC	726
OY	718	CTGCAGAGAGCATCGCTGGAATGACAGCAACCCCGCTGCGCGCGCATCTAC	777
Db	727	CTGCAGAGAGCATCGCTGGAATGACAGCAACCCCGCATCCCGTGGGCGCATCTAC	786
OY	778	AAGCGGTGATCATCTGGGCTGAAACAAGATCGTGCGATGTACAGCCCGTGAGCATC	837
Db	787	AAGCGGTGATCATCTGGGCTGAAACAAGATCGTGCGATGTACAGCCCGTGAGCATC	846
OY	838	CTGCATATCCGCGCAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG	897
Db	847	CTGCATATCAAGCAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG	906
OY	898	ACCCTGCGCGCGAGCAGGCGCACCCAGAGCGTGAAGAACTGGATGACCGAGACCTGCTG	957
Db	907	ACCCTGCGCGCGAGCAGGCGCACCCAGAGCGTGAAGAACTGGATGACCGAGACCTGCTG	966
OY	958	GTCGAGAAAGCGCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGGCGCCACC	1017
Db	967	GTCGAGAAAGCGCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGGCGCCACC	1026
OY	1018	CTGAGAGAGATGATGACCGCTGCCAGGGCGTGGGCGGCCCGGCAACAAGGCCGCGTG	1077
Db	1027	CTGAGAGAGATGATGACCGCTGCCAGGGCGTGGGCGGCCCGGCAACAAGGCCGCGTG	1086
OY	1078	CTGGCCGAGCGGATGAGCCAGGCCCAACAGCGTGAACATCATGTGCAAGAAGCAACTTC	1137
Db	1087	CTGGCCGAGCGGATGAGCCAGGCCCAACAC--AGCGTGTATGTACAGAAGCAACTTC	1143
OY	1138	AAGGGCCCCCGGCGCAACGTCAAGTGTTCCTTCAACTGCGGCAAGAGAGGCCCATCGCCAA	1197
Db	1144	AAGGGCCCCCGGCGCATGTGTCAAGTGTTCCTTCAACTGCGGCAAGAGAGGCCCATCGCCCG	1203

QY	1027	CTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGGCCCAAGCCAAGGCCCGCGTG	1086
Dd	1027	CTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGGCCCAAGCCAAGGCCCGCGTG	1086
QY	1078	CTGGCCGAGGCGCATGAGCCAGGCCAACACGCCTGAACATCATGTGCAAGAAGCAAATTTC	1137
Dd	1087	CTGGCCGAGGCGCATGAGCCAGGCCAACACC---AGCGTAGATGATGCAGAAGACAATTTC	1143
QY	1138	AAGGGCCCCCGGCGCAACGTTCAAGTGCCTTCAACTGGCGGCAAGAGGGGCCCATCGCCAAAG	1197
Dd	1144	AAGGGCCCCCGGCGCATCGTCGTAAGTGCTTTCAACTGCGGCAAGAGAGGGGCCCATCGCCCGC	1203



QY 1378 CCCCCCGGAGAGCTTCGCTTCGAGAGACCAACCCCGCCCCCAAGCAGAGCCCAAG 1437  
DB 1366 CCCCCCGGAGAGCTTCGCTTCGAGAGACCAACCCCGCCCCCAAGCAGAGCCCAAG 1425  
QY 1438 GACCGGAGAGCCCTACCGGAGAGCCCTGACCGCCCTGCGCAGCCTGTTGCGCAGCGGCC 1497  
DB 1426 G-----ACCGGAGAGCCCTGACCAAGCCTGAAGAGCCTGTTGCGCAACGAGCC 1473  
QY 1498 CTGAGCCAGTAA 1509  
DB 1474 CTGAGCCAGAAA 1485

## RESULT 11

US-10-190-435-19  
; Sequence 19, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 4419  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagRTmutatRevNef\_C  
US-10-190-435-19

Query Match 84.5%; Score 1275.2; DB 15; Length 4419;  
Best Local Similarity 92.5%; Pred. No. 8.7e-267;  
Matches 1398; Conservative 0; Mismatches 78; Indels 36; Gaps 4;

QY 1 ATGGGCGCCGCGCGCAGCATCTGCGCGCGGAGAGCTGGAACAAGTGGAGAAGATCCGC 60  
DB 7 ATGGGCGCCGCGCGCAGCATCTGCGCGCGGAGAGCTGGAACAAGTGGAGAAGATCCGC 66  
QY 61 CTGCGCCCGCGCGGCAAGAAGCACTAATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
DB 67 CTGCGCCCGCGCGGCAAGAAGTGTATGATGAAGCACTGTGTGGCCAGCCGCGAG 126  
QY 121 CTGAGGGCTTTCGCTTGAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCATC 180  
DB 127 CTGAGAGAGTTCCGCTTGAACCCCGGCTGTGAGACCGAGGCTGCAAGCATC 186  
QY 181 ATGAAGCAGCTGACGCGCGCTGCAACCGGCAACGAGAGCTGCGCAGCCTGTACAAC 240  
DB 187 ATCGGCAAGCTGACGCGCGCTGCAACCGGCAACGAGAGCTGAGAGCCTGTACAAC 246  
QY 241 ACCGTGGCCACCTGTACTGCTGCAACCGGCAAGTCCGCAACCAAGAGGCC 300  
DB 247 ACCGTGGCCACCTGTACTGCTGCAACGAGAGTCCGCAACCAAGAGGCC 306  
QY 301 CTGGAACAAGTCAGAGAGAGAGCAAGCAAGTCCCAAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAACAAGTCAGAGAGAGAGCAAGCAAGTCCCAAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGGAC--GCAAGTGAAGCAGAACTAACCCTATGTCAGAACTGCAAGGCGCAGATG 417  
DB 367 GCGGACAGGCAAGTGAAGCAGAACTAACCCTATGTCAGAACTGCAAGGCGCAGATG 426  
QY 418 GTGACCAAGGCGCATCAAGCCCGCAGACCTGAAGCCTGGGTGAAGGTGATCGAGAGAG 477  
DB 427 GTGACCAAGGCGCATCAAGCCCGCAGACCTGAAGCCTGGGTGAAGGTGATCGAGAGAG 486

QY 478 GCCTTCAAGCCCGGAGGTGATCCCAATGTTCAACCGCCTGAGCGAGGCGCCACCCCCAG 537  
DB 487 GCCTTCAAGCCCGGAGGTGATCCCAATGTTCAACCGCCTGAGCGAGGCGCCACCCCCAG 546  
QY 538 GACCTGAACAAGATGTTGAACAACCGTGGGCGGCGCACAGGCGCGCAATGATGCTGAAG 597  
DB 547 GACCTGAACAAGATGTTGAACAACCGTGGGCGGCGCACAGGCGCGCAATGATGCTGAAG 606  
QY 598 GACACCATCAACGAGAGAGCGCGCGAGTGGGACCGCTGCAACCGCTGCAAGCGCGGCC 657  
DB 607 GACACCATCAACGAGAGAGCGCGCGAGTGGGACCGCTGCAACCGCTGCAAGCGCGGCC 666  
QY 658 GTGCCCCCGGCGCAGATGCGCGCACCCCGCGCGCAGCGACATCGCGCGGCCACAGACA 717  
DB 667 ATCGCCCCCGGCGCAGATGCGCGCACCCCGCGCGCAGCGACATCGCGCGGCCACAGACC 726  
QY 718 CTGCAAGAGCAGATCGCCTGATGACCAAGCAACCCCGCGTGGCGGCGACATCTAC 777  
DB 727 CTGCAAGAGCAGATCGCCTGATGACCAAGCAACCCCGCGTGGCGGCGACATCTAC 786  
QY 778 AAGCGGTGATCATCTGGGCTGAAACAAGATCGTGGGATGTACACCCCGGTAGCATC 837  
DB 787 AAGCGGTGATCATCTGGGCTGAAACAAGATCGTGGGATGTACACCCCGGTAGCATC 846  
QY 838 CTGACATCCGCGCAGGCGCCCAAGAGACCTTCCGCGACTACGTGACCGCTTCTTCAAG 897  
DB 847 CTGACATCAAGCAGGCGCCCAAGAGACCTTCCGCGACTACGTGACCGCTTCTTCAAG 906  
QY 898 ACCCTGCGCGCGCGAGCAAGGCCACCAAGACGTGAAGAAGTGAAGACCGAGACCTGCTG 957  
DB 907 ACCCTGCGCGCGCGAGCAAGGCCACCAAGAGTGAAGAAGTGAAGACCGAGACCTGCTG 966  
QY 958 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCCCGCGCAC 1017  
DB 967 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCCCGCGCAC 1026  
QY 1018 CTGAGAGAGATGATGACCGCCTGCGCAGGCGCTGGCGGCGCCCGCGCACAGGCGCGCTG 1077  
DB 1027 CTGAGAGAGATGATGACCGCCTGCGCAGGCGCTGGCGGCGCCCGCGCACAGGCGCGCTG 1086  
QY 1078 CTGCGCGGAGGCGGATGAGCCAGGCGCAACAGCGTGAACATCATGATGAGAGAGCACTTC 1137  
DB 1087 CTGCGCGGAGGCGGATGAGCCAGGCGCAAC--AGCGTATGATGAGAGAGCACTTC 1143  
QY 1138 AAGGCGCCCGCGCGCAAGCTCAAGTGTCTCACTGCGCGCAAGAGAGGCCACATCGCGCAAG 1197  
DB 1144 AAGGCGCCCGCGCGCAAGTGTCTCAAGTGTCTCACTGCGCGCAAGAGAGGCCACATCGCGCGC 1203  
QY 1198 AACTGCGCGCGCCCGCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGAGGCCACCATG 1257  
DB 1204 AACTGCGCGCGCCCGCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGAGGCCACCATG 1263  
QY 1258 AAGGCTGCAACGAGCGCGCAAGCTTCTGCGCAAGATCTGCGCGCAAGAGAGGCC 1317  
DB 1264 AAGGCTGCAACGAGCGCGCAAGCTTCTGCGCAAGATCTGCGCGCAAGAGAGGCC 1323  
QY 1318 CGCCCGCGCAACTTCTGCAAGAACGCGAGCGCGCGCGCCCAAGCTGCGCGCACCGCC 1377  
DB 1324 CGCCCGCGCAACTTCTGCAAGAACGCG--CGCGAGCGCGCACCGCC 1365  
QY 1378 CCCCCCGGAGAGCTTCGCTTCGAGAGACCAACCCCGCCCCCAAGCAGAGCCCAAG 1437  
DB 1366 CCCCCCGGAGAGCTTCGCTTCGAGAGACCAACCCCGCCCCCAAGCAGAGCCCAAG 1425  
QY 1438 GACCGGAGCCCTAACCGGAGCGCGCGTGAAGCCTGCGCAGCCTGTTGCGCAGCGGCC 1497  
DB 1426 G-----ACCGGAGAGCCCTGACCAAGCCTGAAGAGCCTGTTGCGCAACGAGCC 1473  
QY 1498 CTGAGCCAGTAA 1509  
DB 1474 CTGAGCCAGAAA 1485

```
RESULT 12
US-10-190-305A-14
; Sequence 14, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 4419
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GagRtmultatRevNef_C
US-10-190-305A-14
```

```
Query Match      84.5%; Score 1275.2; DB 16; Length 4419;
Best Local Similarity 92.5%; Pred. No. 8.7e-267;
Matches 1398; Conservative 0; Mismatches 78; Indels 36; Gaps 4;
```

```
QY 1 ATGGGCGCCCGCGCGCATCTCGCGGCGGAGAGCTGGACAAGTGGAGAAGATCCGC 60
Db 7 ATGGGCGCCCGCGCGCATCTCGCGGCGGAGAGCTGGACGCGCTGGAGCGCATCCGC 66
QY 61 CTGCGCCCCGCGCGCAAGAAGCACTACATGCTGAAGCACTGTGTGGGCCAGCCCGAG 120
Db 67 CTGCGCCCCGCGCGCAAGAAGTGTACATGATGAAGCACTGTGTGGGCCAGCCCGAG 126
QY 121 CTGAGGGGCTTCGCGCTGAACCCCGGCTGTGAGACCGCGAGGGCTGCAAGCATC 180
Db 127 CTGAGAGAGTTCGCGCTGAACCCCGGCTGTGAGACCAAGAGGGCTGCAAGCATC 186
QY 181 ATGAAGCAGCTGACCGCCGCTGACAGCCGACCGAGGAGCTGCGCAGCTTACAAC 240
Db 187 ATCGCGCAGCTGACCGCCGCTGACAGCCGACCGAGGAGGAGCTGAAGAGCTTCAAC 246
QY 241 ACCGTGGCCACCTGTACTGCTGTCACCGCGCATCGAGTCCGCGACCAAGAGGCC 300
Db 247 ACCGTGGCCACCTGTACTGCTGTCACGAGAATCGAGTCCGCGACCAAGAGGCC 306
QY 301 CTGGAACAAGATCGAGGAGGAGAGCAAGTCCGACGAGAAGACCAAGAGGCCAAGAG 360
Db 307 CTGGAACAAGATCGAGGAGGAGAGCAAGTCCGACGAGAAGATCCAGAGGCCAAGAG 366
QY 361 GCGGAC--GGCAAGTGAAGCCAGAACTACCCCATCTGTGAGAACTGCAAGGGCCAGATG 417
Db 367 GCGGACAAAGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGCAAGGGCCAGATG 426
QY 418 GTGCAACGAGGCAATCAAGCCCGGACCCCTGAACGCTGGTGAAGGTGATCGAGAGAAG 477
Db 427 GTGCAACGAGGCAATCAAGCCCGGACCCCTGAACGCTGGTGAAGGTGATCGAGAGAAG 486
QY 478 GCCTTCAGCCCGGAGGTGATCCCATATGTTCAACGCGCTGAGCGAGGGGCCACCCCCAG 537
Db 487 GCCTTCAGCCCGGAGGTGATCCCATATGTTCAACGCGCTGAGCGAGGGGCCACCCCCAG 546
QY 538 GACCTGAACACGATGTTGAACACCGGTGGCGGCGGACCAAGGCCCATGACATGCTGAAG 597
Db 547 GACCTGAACACGATGTTGAACACCGGTGGCGGCGGACCAAGGCCCATGACATGCTGAAG 606
QY 598 GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGTGCAGGCGGGCCCC 657
Db 607 GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGTGCAGGCGGGCCCC 666
```

```
QY 658 GTGCCCCCGCGCAGATGCGCGACCCCGCGGACGACATCGCGGCGCACCAAGCAC 717
Db 667 ATGCCCCCGCGCAGATGCGCGACCCCGCGGACGACATCGCGGCGCACCAAGCAC 726
QY 718 CTGCAAGAGCAGATCGCCTGTGATGACCAAGACCCCGCGTGGCGGCGCATCTAC 777
Db 727 CTGCAAGAGCAGATCGCCTGTGATGACCAAGACCCCGCGTGGCGGCGCATCTAC 786
QY 778 AAGCGGTGATCATCTGTGGGCTGAACAAGATCTGCGGATGACAGCCCGTGAAGCATC 837
Db 787 AAGCGGTGATCATCTGTGGGCTGAACAAGATCTGCGGATGACAGCCCGTGAAGCATC 846
QY 838 CTGACATCGCGCAGGCGCCCAAGAGGCGCTTCCGACTACGTGAGCCGCTTCTCAAG 897
Db 847 CTGACATCAAGCAGGCGCCCAAGAGGCGCTTCCGACTACGTGAGCCGCTTCTCAAG 906
QY 898 ACCCTGCGCGCGAGCAGGCGCACCCAGACCTGAAGAACTGATGACCGACCTGCTG 957
Db 907 ACCCTGCGCGCGAGCAGAGCACCCAGAGGTGAAGAACTGATGACCGACCTGCTG 966
QY 958 GTGCAAGACGCCAACCCCGACTGCAAGACCACTTGGCGGCTTGGCGCGCGCGCAC 1017
Db 967 GTGCAAGACGCCAACCCCGACTGCAAGACCACTTGGCGGCTTGGCGCGCGCGCAC 1026
QY 1018 CTGAGAGAGATGATGACCGCTGACAGGCGGTGGCGGCGCGCGCACCAAGCGCGTG 1077
Db 1027 CTGAGAGAGATGATGACCGCTGACAGGCGGTGGCGGCGCGCGCACCAAGCGCGTG 1086
QY 1078 CTGCGCGAGGCGATGAGCGCAGCGCAACAGCGTGAACATCATGATGCAAGAACACTTC 1137
Db 1087 CTGCGCGAGGCGATGAGCGCAGCGCAACAGCGTGAACATCATGATGCAAGAACACTTC 1143
QY 1138 AAGGCCCCCGCGCAACGTCAAGTCTTCACTGCGGCAAGAGGCGCAATCGCCAA 1197
Db 1144 AAGGCCCCCGCGCATCTGCAAGTCTTCACTGCGGCAAGAGGCGCAATCGCCCGC 1203
QY 1198 AACTGCGCGCGCGCGCAAGAGGCGTGTGAAGTGCAGCAAGAGGCGCACAGATG 1257
Db 1204 AACTGCGCGCGCGCGCAAGAGGCGTGTGAAGTGCAGCAAGAGGCGCACAGATG 1263
QY 1258 AAGACTGCAACGAGCGCGCAACTTCTGGGCAAGATCTGGCCCAAGCAAGGGC 1317
Db 1264 AAGACTGCAACGAGCGCGCAACTTCTGGGCAAGATCTGGCCCAAGCAAGGGC 1323
QY 1318 CGCCCCGGCACTTCTGTGAAGAACCGCAGAGCGCGCGCCCAACCTGTCCACCGCC 1377
Db 1324 CGCCCCGGCACTTCTGTGAAGAGCG-----CGCGAGCCACCGCC 1365
QY 1378 CCCCCCGCGAGAGCTTCGCTTGAAGAGACACCCCGCGCCCAAGAGAGCCCAAG 1437
Db 1366 CCCCCCGCGAGAGCTTCGCTTGAAGAGACACCCCGCGCGCAAGAGAGAGCAAG 1425
QY 1438 GACCGGAGCGCTTACCGGAGCGCGCTGACCGCGCTGCGAGCTGTTGGAGCGCGCC 1497
Db 1426 G-----ACCGGAGAGCGCTGACCAAGCTGAAGAGCTGTTGGCAAGAGCC 1473
QY 1498 CTGAGCCAGTAA 1509
Db 1474 CTGAGCCAGAAA 1485
```

```
RESULT 13
US-10-190-435-35
; Sequence 35, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
```

```
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 4483
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; US-10-190-435-35
```

```
Query Match      84.5%; Score 1275.2; DB 15; Length 4483;
Best Local Similarity 92.5%; Pred. No. 8.7e-267;
Matches 1398; Conservative 0; Mismatches 78; Indels 36; Gaps 4;

QY      1 ATGGGCGCGCGCGCGCATCTCGCGCGCGAGAAAGCTGGACAAGTGGAGAGATCCGC 60
DB      3005 ATGGGCGCGCGCGCGCATCTCGCGCGCGAGCAAGCTGGACGCTGGGAGCGCATCCGC 3064

QY      61 CTGCGCGCGCGCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCCAGCGCGAG 120
DB      3065 CTGCGCGCGCGCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGGCCAGCGCGAG 3124

QY      121 CTGAGAGGCTTCCGCTTGAACCCCGGCTGTGAGACCGCGAGGGCTGCAAGCAGATC 180
DB      3125 CTGAGAGAGTTCGCTTGAACCCCGGCTGTGAGACCGAGGGCTGCAAGCAGATC 3184

QY      181 ATGAAGCAGCTGACCGCGCGCGCGCGCGAGACCGGACGAGAGAGCTGGCAGCTGTCAAC 240
DB      3185 ATCCGCGAGCTGACCGCGCGCGCGCGCGAGACCGGACGAGAGAGCTGTCAAC 3244

QY      241 ACCGTGGCCACCTGTACTGCTGTCACGCGGCAATGAGTCCGCGCAACCAAGAGAGGCC 300
DB      3245 ACCGTGGCCACCTGTACTGCTGTCACGAGAGATGAGTCCGCGCAACCAAGAGAGGCC 3304

QY      301 CTGGAACAAGATCGAGAGAGAGAGAGAGCAAGTCCGACAGAGAGAGAGAGAGAGAG 360
DB      3305 CTGGAACAAGATCGAGAGAGAGAGAGAGAGCAAGTCCGACAGAGAGAGAGAGAGAGAG 3364

QY      361 GCCGAC--GGCAAGGTGAGCCAGAACTACCCCATGCTGAGAACTGCAAGGCGCAGATG 417
DB      3365 GCCGACAAGGCAAGGTGAGCCAGAACTACCCCATGCTGAGAACTGCAAGGCGCAGATG 3424

QY      418 GTGCAACGAGGCAATCAGCCCCCGGACCCCTGAACGCTGGGTGAAGGTATCGAGAGAA 477
DB      3425 GTGCAACGAGGCAATCAGCCCCCGGACCCCTGAACGCTGGGTGAAGGTATCGAGAGAA 3484

QY      478 GCCTTCAGCCCCCGAGGTGATCCCAATGTTCAACGCGCTGAGCGAGGGCGCCACCCCC 537
DB      3485 GCCTTCAGCCCCCGAGGTGATCCCAATGTTCAACGCGCTGAGCGAGGGCGCCACCCCCAG 3544

QY      538 GACCTGAACACGATGTTGAACACCCGTGGGGCGGCGCAACGAGCGCCCATGCAATGCTGA 597
DB      3545 GACCTGAACACGATGTTGAACACCCGTGGGGCGGCGCAACGAGCGCCCATGCAATGCTGA 3604

QY      598 GACACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
DB      3605 GACACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3664

QY      658 GTGGCCCCCGCGCGAGATGCGCGAGACCCCGCGGCGAGAGAGATCGCGCGCGCCAC 717
DB      3665 ATCGCCCCCGCGCGAGATGCGCGAGACCCCGCGGCGAGAGATCGCGCGCGCCAC 3724

QY      718 CTGCAAGAGAGAGATGCGCTGATGACCAAGCAACCCCGCTGCGCTGGGCGACATCTAC 777
DB      3725 CTGCAAGAGAGAGATGCGCTGATGACCAAGCAACCCCGCTGCGCTGGGCGACATCTAC 3784

QY      778 AAGCGGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC 837
DB      3785 AAGCGGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC 3844
```

```
QY      838 CTGACATCCGCCAGGGCCCCCAAGAGCCCTTCCGCACTACGTGAGCCGCTTCTCAAG 897
DB      3845 CTGACATCAAGCAGGGCCCCCAAGAGCCCTTCCGCACTACGTGAGCCGCTTCTCAAG 3904

QY      898 ACCCTCGCGCGCGAGCAGGCCACCCAGAGAGCTGAAGACTGATGACCGAGACCCTGCTG 957
DB      3905 ACCCTCGCGCGCGAGCAGAGCACCAGAGAGTGAAGACTGATGACCGAGACCCTGCTG 3964

QY      958 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTCCGCGCGCTTCCGCGCGCGCGCAC 1017
DB      3965 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTCCGCGCGCTTCCGCGCGCGCGCAC 4024

QY      1018 CTGAGAGAGATGATGACCGCTGCGCAGGGCGTGGGGCGCGCGCGCACAGCGCGGCTG 1077
DB      4025 CTGAGAGAGATGATGACCGCTGCGCAGGGCGTGGGGCGCGCGCGCACAGCGCGGCTG 4084

QY      1078 CTGCGCGGAGCGGATGAGCGGCAAGCGCTGAACATCATGATGACAGAGACCACTTC 1137
DB      4085 CTGCGCGGAGCGGATGAGCGGCAAGCGCTGAACATCATGATGACAGAGACCACTTC 4141

QY      1138 AAGGCCCCCGCGCGCAAGTCAAGTCTTCACTGCGGCAAGAGAGGCCACATCGCCAA 1197
DB      4142 AAGGCCCCCGCGCGCATGTCAGTCTTCACTGCGGCAAGAGAGGCCACATCGCCCGC 4201

QY      1198 AACTGCGCGCGCGCGCGCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGAGGCCAC 1257
DB      4202 AACTGCGCGCGCGCGCGCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGAGGCCAC 4261

QY      1258 AAGGACTGCAACCGAGCGCGCAAGGCTTCTGCGGCAAGATCTGGCGCAAGAGAGGC 1317
DB      4262 AAGGACTGCAACCGAGCGCGCAAGGCTTCTGCGGCAAGATCTGGCGCAAGAGAGGC 4321

QY      1318 CGCGCGCGCAACTTCTGTCAGAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
DB      4322 CGCGCGCGCAACTTCTGTCAGAGCGG-----CCCGAGCCACCGCC 4363

QY      1378 CCGCGCGCGGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
DB      4364 CCGCGCGCGGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4423

QY      1438 GACCGCGAGCCCTAACCGGAGCGGCTGACCGCGCTGCGGAGCTGTTCGGCAGCGGCGCC 1497
DB      4424 G-----ACCGGAGAGCCCTGACAGAGCTGAAGAGCTGTTCGGCAACGACCC 4471

QY      1498 CTGAGCCAGTAA 1509
DB      4472 CTGAGCCCAATAA 4483

RESULT 14
US-10-190-435-34
; Sequence 34, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 4606
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
```

OTHER INFORMATION: gp160mod.TV1.dv1-gagmod.BW965  
US-10-190-435-34

Query Match 84.5%; Score 1275.2; DB 15; Length 4606;  
Best Local Similarity 92.5%; Pred. No. 8.7e-267;  
Matches 1398; Conservative 0; Mismatches 78; Indels 36; Gaps 4;

```
QY 1 ATGGCGCGCGCGCCAGCATCTCGCGCGGAGAGCTGGACAGTGGAGAGATCCGC 60
Db 3128 ATGGCGCGCGCGCCAGCATCTCGCGCGGAGAGCTGGAGCGCATCCGC 3187
QY 61 CTGCGCGCGCGCGCAAGAGCACTAGTGTGAAGCACTGTGTGGGCCAGCCGCGAG 120
Db 3188 CTGCGCGCGCGCGCAAGAGTGTAGTGAAGCACTGTGTGGGCCAGCCGCGAG 3247
QY 121 CTGAGGCGTTGCGCTGTAACCCCGCGCTGTGAAGCCCGAGGGGTGCAAGCATC 180
Db 3248 CTGAGAGATTGCGCTGTAACCCCGCGCTGTGAAGCCAGGGGTGCAAGCATC 3307
QY 181 ATGAAGCAGCTGACGCCCGCTGACAGCCGACCGAGAGCTGCGCAGCTGTACAAC 240
Db 3308 ATCCGCGAGCTGACCCCGCGCTGACAGCCGACCGAGAGAGCTGAAGAGCTGTACAAC 3367
QY 241 ACCGTGCGCACCCCTGTACTGCGTGACGCGCGCATGAGTCCGCGACCAAGAGGCGC 300
Db 3368 ACCGTGCGCACCCCTGTACTGCGTGACAGAGAGATGAGTCCGCGACCAAGAGGCGC 3427
QY 301 CTGACAAAGATCGAGAGAGAGAGCAAGTCCAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 3428 CTGACAAAGATCGAGAGAGAGAGAGCAAGTCCAGAGAGAGAGATCCAGAGAGAGAGAG 3487
QY 361 GCCGAC---GGCAAGGTGAGCCAGAGACTACCCCATGTCAGAACTGAGAGGCGCAGATG 417
Db 3488 GCCGACAAAGGTGAGAGAGAGAGAGCAAGTACCCCATGTCAGAACTGAGAGGCGCAGATG 3547
QY 418 GTGACACGAGGCGCATACGCCCCCGACCTGTAAGCGCTGAGAGTGTGAGAGAGAG 477
Db 3548 GTGACACGAGGCGCATACGCCCCCGACCTGTAAGCGCTGAGAGTGTGAGAGAGAG 3607
QY 478 GCCTTCAGCCCCGAGGTGATCCCACTGTTCAACGCGCTGAGCGAGGCGCCACCCCGCAG 537
Db 3608 GCCTTCAGCCCCGAGGTGATCCCACTGTTCAACGCGCTGAGCGAGGCGCCACCCCGCAG 3667
QY 538 GACCTGAACACGATGTGAACACCGTGGGCGGCGCACAGGCGCGCATGAGATGCTGAAG 597
Db 3668 GACCTGAACACGATGTGAACACCGTGGGCGGCGCACAGGCGCGCATGAGATGCTGAAG 3727
QY 598 GACACCATCAACGAGAGAGGCGCGAGTGGAGCGCGCTGACCCCGTGCAGCGCGCGCC 657
Db 3728 GACACCATCAACGAGAGAGGCGCGAGTGGAGCGCGCTGACCCCGTGCAGCGCGCGCC 3787
QY 658 GTGGCCCCCGCGCAGATGCGCGACCGCGCGCGCGCAGCATGCGCGCGCGCACAGCAC 717
Db 3788 ATGCGCCCCGCGCAGATGCGCGACCGCGCGCGCGCAGCATGCGCGCGCGCACAGCAC 3847
QY 718 CTGCAAGAGCAGATGCGCTGATGACAGCAACCCCGCGTGGCGCGCATCTAC 777
Db 3848 CTGCAAGAGCAGATGCGCTGATGACAGCAACCCCGCGTGGCGCGCATCTAC 3907
QY 778 AAGCGGTGATCATCTGGGCGCTGAACAAGATGTCGAGATGACAGCCCGTGAAGCATC 837
Db 3908 AAGCGGTGATCATCTGGGCGCTGAACAAGATGTCGAGATGACAGCCCGTGAAGCATC 3967
QY 838 CTGGAACATCCGCGAGGCGCGCAAGAGAGCCCTTCGCGAGTACGTGAGACCGCTTCAAG 897
Db 3968 CTGGAACATCAAGGAGGCGCGCAAGAGAGCCCTTCGCGAGTACGTGAGACCGCTTCAAG 4027
QY 898 ACCCTGCGCGCGAGCAGGCGCACCCAGAGAGTGAAGAACTGATGACCGAGACCTGTCTG 957
Db 4028 ACCCTGCGCGCGAGCAGGCGCACCCAGAGAGTGAAGAACTGATGACCGAGACCTGTCTG 4087
QY 958 GTGCAAGAGCGCAACCCCGCATGCAAGACCATCTGCGCGCTCTCGCGCGCGCGCGCAC 1017
Db 3137 ATGGCGCGCGCGCGCCAGCATCTCGCGCGGAGAGAGTGGAGAGAGATCCGC 3196
```

```
Db 4088 GTGCAAGAGCGCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCGCGCGCGCAGC 4147
QY 1018 CTGAGAGAGATGATGACCCGCTGCGCAGGCGGTGGGCGGCGCGCGCACAAAGCGCGGTG 1077
Db 4148 CTGAGAGAGATGATGACCCGCTGCGCAGGCGGTGGGCGGCGCGCGCACAAAGCGCGGTG 4207
QY 1078 CTGGCGGAGGCGGATGAGCCAGGCGCAACAGCGTGAACATCATGATGAGAGAGCACTTC 1137
Db 4208 CTGGCGGAGGCGGATGAGCCAGGCGCAACACC---AGCGTATGATGAGAGAGCACTTC 4264
QY 1138 AAGGCGCGCGCGCGCAACGTCAAGTGTCTTCAACTGCGGCGCAAGAGGCGCACATGCGCAAG 1197
Db 4265 AAGGCGCGCGCGCGCATGCGTCAAGTGTCTTCAACTGCGGCGCAAGAGGCGCACATGCGCGC 4324
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGCGCTGCTGAAGTGCAGGAGAGAGGCGCACAGATG 1257
Db 4325 AACTGCGCGCGCGCGCGCGCAAGAGGCGCTGCTGAAGTGCAGGAGAGAGGCGCACAGATG 4384
QY 1258 AAGGACTGCAACGAGCGCGCAGGCGCAACTTCTGCGGCAAGATGCGCCAGCGCAAGGCGC 1317
Db 4385 AAGGACTGCAACGAGCGCGCAGGCGCAACTTCTGCGGCAAGATGCGCCAGCGCAAGGCGC 4444
QY 1318 CGCGCGCGCAACTTCTGTCAGAACCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1377
Db 4445 CGCGCGCGCAACTTCTGTCAGAGCGG-----CGCGAGCGCGCGCGCGCGCGCGCGCGC 4486
QY 1378 CCCCCCGCGCAGAGCTTCCGCTTGCAGAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGC 1437
Db 4487 CCCCCCGCGCAGAGCTTCCGCTTGCAGAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGC 4546
QY 1438 GACCGCGAGCGCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1497
Db 4547 G-----ACCGCGAGACCTGACCAAGCTGTAAGAGCGCTGTGCGCAAGCAGACCGC 4594
QY 1498 CTGAGCCAGTAA 1509
Db 4595 CTGAGCCAAATAA 4606
```

RESULT 15  
US-10-190-435-36  
; Sequence 36, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDR, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 36  
; LENGTH: 4615  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: gp160mod.TV1.dv2-gagmod.BW965  
US-10-190-435-36

Query Match 84.5%; Score 1275.2; DB 15; Length 4615;  
Best Local Similarity 92.5%; Pred. No. 8.7e-267;  
Matches 1398; Conservative 0; Mismatches 78; Indels 36; Gaps 4;

```
QY 1 ATGGCGCGCGCGCGCCAGCATCTCGCGCGGAGAGAGTGGAGAGAGATCCGC 60
Db 3137 ATGGCGCGCGCGCGCCAGCATCTCGCGCGGAGAGAGTGGAGAGAGATCCGC 3196
```

QY	61	CTGGCCCCCGGCGGCAAGAACCTACATGCTGAAGCACCTGTGTGGGCCAGCCGGCAG	120
Db	3197	CTGGCCCCCGGCGGCAAGAGTGCTACATGATGAAGCACCTGTGTGGGCCAGCCGGCAG	3256
QY	121	CTGGAGGGGCTTGCCTCTGAACCCCGGGCCTGTGTGAGACCGCCGAGGGCTGCAAGCATC	180
Db	3257	CTGGAGAAAGTTCGCTCTGAACCCCGGGCCTGTGTGAGACCAAGGAGGGCTGCAAGCATC	3316
QY	181	ATGAAGCAGCTGCAGCCCGCCCTGCAGACCCGGCACCGAGAGCTGCGCAGCCTGTACAAC	240
Db	3317	ATCCGCCAGCTGCACCCCGCCCTGCAGACCCGGCAGCGAGAGCTGAAGAGCCTGTTCAAC	3376
QY	241	ACCGTGGCCACCCCTGTACTGCGTGCA CGCCGGCATGAGGTCCGCGACCAAGAGAGCC	300
Db	3377	ACCGTGGCCACCCCTGTACTGCGTGCA CGAAGAGATGAGGTCCGCGACCAAGAGAGCC	3436
QY	301	CTGGAACAAGATCGAGGAGAGAGCAACAAGTCCAGCAGAAAGACCAGCAGGCCAAGAG	360
Db	3437	CTGGAACAAGATCGAGGAGAGAGCAACAAGTCCAGCAGAAAGATCCAGCAGGCCAGAGCC	3496
QY	361	GCCGCA--GGCAAGGTAGGCCCAACTACCCCATCTGTGAGAACTTGACGGCCAGATG	417
Db	3497	GCCGCAAGGGCAAGGTAGGCCCAAGAACTACCCCATCTGTGAGAACTTGACGGGCCAGATG	3556
QY	418	GTCGACACGAGGCCATCAGCCCCCGGCACCTTGAAAGCCTGGGTGAAGGTATCGAGAGAA	477
Db	3557	GTCGACACGAGGCCATCAGCCCCCGGCACCTTGAAAGCCTGGGTGAAGGTATCGAGAGAA	3616
QY	478	GCCTTCAGCCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGGCCACCCCCAG	537
Db	3617	GCCTTCAGCCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGGCCACCCCCAG	3676
QY	538	GACCTGAACAAGATGTTGAACACCGTGGGCGGCCACAGGCCGCCATGCAATGCTGAAG	597
Db	3677	GACCTGAACAAGATGTTGAACACCGTGGGCGGCCACAGGCCGCCATGCAATGCTGAAG	3736
QY	598	GACACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCCTGCACCCCGTGACGGCGGCC	657
Db	3737	GACACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCCTGCACCCCGTGACGGCGGCC	3796
QY	658	GTCGCCCCCGGCGAGATGCGCGCACCCCGCGGCGAGCAATCGCCGGCGCACCAAGCAC	717
Db	3797	ATCGCCCCCGGCGAGATGCGCGCACCCCGCGGCGAGCAATCGCCGGCGCACCAAGCAC	3856
QY	718	CTGCAAGAGCAGATCGCCTGATGACCAAGCAACCCCGCGTGGCGCGACATCTAC	777
Db	3857	CTGCAAGAGCAGATCGCCTGATGACCAAGCAACCCCGCATCCCGTGGCGACATCTAC	3916
QY	778	AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTAAGCCCGTGAGCATC	837
Db	3917	AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTAAGCCCGTGAGCATC	3976
QY	838	CTGCAATCCGCGAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTAAG	897
Db	3977	CTGCAATCAAGCAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTAAG	4036
QY	898	ACCCTGCGCGCGAGCAGGCGCACCCAGGACGTGAAGAACTGATGACCGAGACCCCTGCTG	957
Db	4037	ACCCTGCGCGCGAGCAGGCGCACCCAGGAGGTGAAGAACTGATGACCGAGACCCCTGCTG	4096
QY	958	GTCGAAACGCGCAACCCCGACTGCAAGAACCATCTGCGCGCTCTCGGCCCGCGCCACC	1017
Db	4097	GTCGAAACGCGCAACCCCGACTGCAAGAACCATCTGCGCGCTCTCGGCCCGCGCCAGC	4156
QY	1018	CTGGAAGAGATGATGACCGCGCTGCGCAGGGCGTGGGGCGGCCCGGCACCAAGGCCGCTG	1077
Db	4157	CTGGAAGAGATGATGACCGCGCTGCGCAGGGCGTGGGGCGGCCCGGCACCAAGGCCGCTG	4216
QY	1078	CTGGCCGAGGCGCATGAGCCAGGCCAACAAGCGGTGAACATCATGATGCAAGAAAGCACTTC	1137
Db	4217	CTGGCCGAGGCGCATGAGCCAGGCCAACAAC---AGCGTATGATGCAAGAAAGCACTTC	4273
QY	1138	AAGGGCCCCCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGGGCCCATCGCCAAG	1197

Db	4274	AAGGGCCCCCGCGCATCGTCAAGTGTCTCACTGCGGCAAGAGGGCCATCGCCCC	4333
QY	1198	AACTGCGCGCCCCCGCAGAAGGGCTGCTGGAAGTGCSCCAAGAGGGCCACCATG	1257
Db	4334	AACTGCGCGCCCCCGCAGAAGGGCTGCTGGAAGTGCSCCAAGAGGGCCACCATG	4393
QY	1258	AAGGACTGCAACCGAGCGGCCAGGCCCACTTCTTGCGCAAGATCTGCCCCAGCCACAAGGGC	1317
Db	4394	AAGGACTGCAACCGAGCGGCCAGGCCCACTTCTTGCGCAAGATCTGCCCCAGCCACAAGGGC	4453
QY	1318	CGCCCCGGCAAATTCTGTGCAAAACCGCAGCGAGCCCGCCGCCACCGTGCACCGCC	1377
Db	4454	CGCCCCGGCAAATTCTGTGCAAGCCG-----CCCCAGCCCAACCGCC	4495
QY	1378	CCCCCGCGCAGAGCTTCCGCTTCGAGGAGACCACCCCGCCCCCAAGCAGAGAGCCCAAG	1437
Db	4496	CCCCCGCGCAGAGCTTCCGCTTCGAGGAGACCACCCCGGCCAGAAACAGAGAGCAAG	4555
QY	1438	GACCGCGAGCCCTAACCGCGAGCCCCCTGACCGCCCTGCGCAGCCTGTTGGCAGCGGCCCC	1497
Db	4556	G-----ACCGCGAGACCCTGACCAGCCTGAAGAGCCTGTTGGCAGCAGACCCC	4603
QY	1498	CTGAGCCAGTAA 1509	
Db	4604	CTGAGCCAGTAA 4615	

Search completed: March 13, 2005, 04:29:43  
Job time : 923.106 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:39:57 ; Search time 178.272 Seconds  
(without alignments)  
5584.002 Million cell updates/sec

Title: US-09-475-704A-4  
Perfect score: 2840  
Sequence: 1 atggcgccgcgcgcagcat.....gcgcgcctgagccagtaa 1509

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1396920 seqs, 329844858 residues

Total number of hits satisfying chosen parameters: 2793840

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US09475704/runat\_10032005\_140224\_14937/app\_query.fasta\_1.3342  
-DB=Published Applications AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62  
-TRANS=human40.cdd -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09475704 @CGN\_1\_1\_466 @runat\_10032005\_140224\_14937  
-NCPU=6 -ICPU=3 -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGIOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCR\_US\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	2661	93.7	502	10	US-09-899-575-22	Sequence 22, Appl
2	2464	86.8	492	16	US-10-332-413-2	Sequence 2, Appl1
3	2457	86.5	492	9	US-09-991-258-5	Sequence 5, Appli
4	2457	86.5	631	16	US-10-332-413-8	Sequence 8, Appl1
5	2440	85.9	492	14	US-10-339-217-109	Sequence 109, App
6	2430	85.6	492	14	US-10-339-217-110	Sequence 110, App
7	2421.5	85.3	491	14	US-10-339-217-140	Sequence 140, App
8	2413.5	85.0	491	14	US-10-339-217-143	Sequence 143, App
9	2387	84.0	508	14	US-10-339-217-105	Sequence 105, App
10	2385.5	84.0	491	14	US-10-339-217-104	Sequence 104, App
11	2375.5	83.6	497	14	US-10-339-217-107	Sequence 107, App
12	2353.5	82.9	487	14	US-10-339-217-106	Sequence 106, App
13	2348	82.7	496	14	US-10-339-217-108	Sequence 108, App
14	2343	82.5	492	10	US-09-899-575-17	Sequence 17, Appl
15	2258	79.5	500	14	US-10-339-217-118	Sequence 18, Appl
16	2247.5	79.1	494	14	US-10-339-217-126	Sequence 126, App
17	2246	79.1	500	14	US-10-223-172A-42	Sequence 42, Appl
18	2246	79.1	500	14	US-10-224-999A-3483	Sequence 3483, Ap
19	2238	78.8	1350	10	US-09-952-060-35	Sequence 35, Appl
20	2238	78.8	1350	16	US-10-380-641-35	Sequence 35, Appl
21	2232	78.6	557	14	US-10-339-217-144	Sequence 144, App
22	2226	78.4	500	16	US-10-325-468-8	Sequence 8, Appl
23	2225.5	78.3	494	14	US-10-339-217-100	Sequence 100, App
24	2223	78.3	500	14	US-10-097-534-32	Sequence 32, Appl
25	2223	78.3	500	14	US-10-059-271-85	Sequence 85, Appl
26	2223	78.3	500	14	US-10-102-622-4	Sequence 4, Appli
27	2223	78.3	500	16	US-10-325-468-34	Sequence 34, Appl
28	2223	78.3	500	17	US-10-844-658-6	Sequence 6, Appli
29	2221	78.2	500	9	US-09-968-355-26	Sequence 26, Appl
30	2221	78.2	583	9	US-09-968-355-17	Sequence 17, Appl
31	2220.5	78.2	499	17	US-10-844-658-2	Sequence 2, Appli
32	2220	78.2	498	16	US-10-325-468-22	Sequence 22, Appl
33	2216.5	78.0	494	14	US-10-339-217-99	Sequence 99, Appl
34	2212.5	77.9	501	14	US-10-339-217-136	Sequence 136, App
35	2212.5	77.9	503	14	US-10-339-217-127	Sequence 127, App
36	2212	77.9	500	14	US-10-339-217-122	Sequence 122, App
37	2212	77.8	500	16	US-10-325-468-17	Sequence 17, Appl
38	2209.5	77.8	494	14	US-10-339-217-101	Sequence 101, App
39	2209.5	77.8	499	15	US-10-296-734-1	Sequence 32, Appl
40	2206.5	77.7	496	14	US-10-339-217-32	Sequence 32, Appl
41	2205	77.6	498	16	US-10-325-468-27	Sequence 27, Appl
42	2199	77.4	500	14	US-10-339-217-121	Sequence 21, App
43	2197	77.4	512	16	US-10-093-953A-25	Sequence 25, Appl
44	2197	77.4	512	16	US-10-093-953A-30	Sequence 30, Appl
45	2193	77.2	493	14	US-10-339-217-111	Sequence 111, App

## ALIGNMENTS

RESULT 1  
US-09-899-575-22  
; Sequence 22, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megede, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; PRIORITY FILING DATE: 2001-07-05  
; PRIORITY FILING DATE: 09/475,704  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus

US-09-899-575-22

## Alignment Scores:

Pred. No.:	7.64e-129	Length:	502
Score:	2661.00	Matches:	502
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	93.70%	Indels:	0
DB:	10	Gaps:	0

US-09-475-704A-4 (1-1509) x US-09-899-575-22 (1-502)

```
QY 1 ATGGGCGCGCGCGCGCATCTCTGGCGCGCGAGAACTGGACAAGTGGGAGAAAGATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyIlySLeuAspLysTrpGluYsIleArg 20
QY 61 CTGGCGCGCGCGCGGAGAGAAAGCACTACATGCTGAAGCACTGTGTGGGCCAGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysIleTyrMetLeuYsHISLeuValTrpAlaSerArgGlu 40
QY 121 CTGGAGGGCTTCGCGCTGAACCCCGCGCTGTGAGAGACCGCGGAGGGCTGCAAGCAGATC 180
Db 41 LeuGluGlyPheAlaLeuAsnProGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60
QY 181 ATGAAGCAGCTGAGACCGCGCGCTGTGAGAGACCGCGAGAGCTGCGCAGCTGTACAAC 240
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGGTGACAGCGCGGCAATCGAGTCCGCGCAACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluValArgAspThrLysGluAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGAGACAAGTCCCGAGCAGAGACCCAGAGGCCAAGAGAG 360
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerGlnGlnLysThrGlnAlaLysGlu 120
QY 361 GCGGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAAGACCTGCAAGGCCAGATGTG 420
Db 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140
QY 421 CACCAAGGCATAGCCCCCGCACCTTGAACGCTGGGTGAAGTGTAGAGAGAAAGGCC 480
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLysAla 160
QY 481 TTCAGCCCCGAGGTGATCCCCCATGTTACCGCGCTGAGCGAGGGCGCCACCCCGCAGAC 540
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 180
QY 541 CTGAACAAGATGTGAACACCGTGGGGCGGCCACAGGCCCGCATGCAAGTGTGAAGGAC 600
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200
QY 601 ACCATCAACGAGAGAGCGCGCGGAGTGGGAGCCGCGCTGCACCCCGTGCAAGCGCGCGCTG 660
Db 201 ThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValGlnAlaGlyProVal 220
QY 661 GCGCGCGCGCGAGATGCGCGAGCCCGCGGCGAGCAGCATGCGCGCGCCAGCAGCACCTG 720
Db 221 AlaProGlyGlnMetArgAspProArgGlySerAspIleAlaGlyAlaThrSerThrLeu 240
QY 721 CAGGAGCAGATCGCTGTGATGACAGCAACCCCGCGTGCGGTGGGCGACACTTACAAG 780
Db 241 GlnGluGlnIleAlaTrpMetThrSerAsnProProValProValGlyAspIleTyrLys 260
QY 781 CGGTGATCATCTCTGGGCTGGAACAAGATCGTGGAGATGTACAGCCCGTGAGCATCTG 840
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIleLeu 280
QY 841 GACATCCGCGCAGGGCGCCAGAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTCAAGACC 900
Db 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 300
QY 901 CTGCGCGCGCAGAGAGCGCCACCGCAGAGCTGAAGAACTGATGACCGAGACCTGTGTG 960
```

```
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrGluThrLeuVal 320
QY 961 CAGAACGCCAACCCCGACTGCAAGAACCATCTCTGCGCGCTCTGCGCCCCGCGCACCTG 1020
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340
QY 1021 GAGGAGATGATGACCGCTGCGCAGGGCGGTGGCGCGCCCCGCGCACAAGGCCCGCTGCTG 1080
Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360
QY 1081 GCGGAGCGCATGAGCCAGCGCAACAGCGTGAACATCATGATTCAGAGAGCAACTTCAAG 1140
Db 361 AlaGluAlaMetSerGlnAlaAsnSerValAsnIleMetMetGlnLysSerAsnPheLys 380
QY 1141 GCGCGCGCGCGCAACGTCAAGTGTCTCACTGCGGCAAGAGAGGCCACATCGCCCAAGAAC 1200
Db 381 GlyProArgArgAsnValLysCysPheAsnCySGlyLysGluGlyHisIleAlaLysAsn 400
QY 1201 TGCCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGGCAAGAGGGCCACAGATGAAG 1260
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420
QY 1261 GACTGCACCGAGCGCGCAAGCTTCTCTGGGCAAGATCTGGCCCCAGCCACAAGGGCGCG 1320
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGCGCACTTCTCTGCAAGACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 441 ProGlyAsnPheLeuGlnAsnArgSerGluProAlaAlaProThrValProThrAlaPro 460
QY 1381 CCGCGCGAGAGCTTCCGCTTGAGAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCG 1440
Db 461 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProLysAsp 480
QY 1441 CGCGAGCCCTAACCGCGAGCGCCCTGACCGCGCTGCGCAGCTGTTCGGCAGCGCGCGCTG 1500
Db 481 ArgGluProTyrArgGluProLeuThrAlaLeuArgSerLeuPheGlySerGlyProLeu 500
QY 1501 AGCCAG 1506
Db 501 SerGln 502
```

## RESULT 2

US-10-332-413-2

; Sequence 2, Application US/10332413

; Publication No. US2004011660A1

; GENERAL INFORMATION:

; APPLICANT: Johnstone, Robert Edward

; APPLICANT: Swanstrom, Ronald Ivar

; APPLICANT: Morris, Lynn

; APPLICANT: Karim, Salim Abdool

; APPLICANT: Williamson, Carolyn

; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H

; TITLE OF INVENTION: Subtype Isolates, Their Genes and Modifications and Derivatives T

; FILE REFERENCE: 45669-281993

; CURRENT APPLICATION NUMBER: US/10/332,413

; CURRENT FILING DATE: 2003-09-12

; PRIOR APPLICATION NUMBER: PCT/IB01/01208

; PRIOR FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/216,995

; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: ZA 2000/3437

; PRIOR FILING DATE: 2000-07-10

; PRIOR APPLICATION NUMBER: ZA 2000/4924

; PRIOR FILING DATE: 2000-09-15

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 492

; TYPE: PRT

; ORGANISM: Human immunodeficiency virus type 1

US-10-332-413-2

Alignment Scores:

Pred. No.: 9.99e-119 Length: 492  
Score: 2464.00 Matches: 466  
Percent Similarity: 95.62% Conservative: 14  
Best Local Similarity: 92.83% Mismatches: 12  
Query Match: 86.76% Indels: 10  
DB: 16 Gaps: 2

US-09-475-704A-4 (1-1509) x US-10-332-413-2 (1-492)

QY 1 ATGGGGCGCCGCGCGCATCTCGCGCGGAGAGCTGGACAAGTGGAGAGAATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCCCGCGCGGAGAGACATATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysHisIleTrpMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGGCTTCGCGCTGAACCCCGCTGTGAGACCGCGGAGGCTGCAAGCATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGCAGCCCGCTGCAGACCGGACCGAGGAGGCTGCGCAGCTGTACAC 240  
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTCTGTACTGCGTGACCGCGGATCGAGGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGAGAGACAAGTCCCAAGAGAGAGAGAGAGAGAGAG 360  
DB 101 LeuAspLysIleGluGluGluGlnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAAGACCTGCAAGGCGCAGATG 420  
DB 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACCAAGGCATCAGCCCCCGGACCCCTGAACGCGCTGGGTGAAGGTATCGAGAGAGGCC 480  
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTTCAACGCGCTGAGCGAGGGGCGCCACCCCGAGGAC 540  
DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGGTGGGCGGCGACAGGCGCGCATGAGATGCTGAAGGAC 600  
DB 181 LeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGGACCGCTGCACCCCGTGCAGGCGCGCGCCGCTG 660  
DB 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgValHisProValHisAlaGlyProIle 220  
QY 661 GCGCGCGCGCAGATGCGCGGACCCCGCGCGGAGCGGACATCGCGCGCGCCACAGACCCCTG 720  
DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGAGCAGATCGCGCTGATGACCAACACCCCGCTGCGCGCGGCGAGCATTTACAAG 780  
DB 241 GlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGGCTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTCG 840  
DB 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIleLeu 280  
QY 841 GACATCCGCGCAGGCGCCCAAGAGAGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900  
DB 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 300  
QY 901 CTGCGCGCGCAGAGAGCGCCACCCAGGACGTGAAGAAGCTGATGACCGAGACCCCTGCTGTG 960

DB 301 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAACGCCAACCCCGACTGCAGAACCATCTGTGGCGCTCTCGGCGCGCGCCACCTG 1020  
DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCGCGCCACAGAGGCCCGCTGCTG 1080  
DB 341 GlnGluMetMetThrAlaCysGlnGlnValGlyGlyProGlyHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGAGCCAGGCGCAACAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140  
DB 361 AlaGluAlaMetSerGlnAlaAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380  
QY 1141 GCGCGCGCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGCGCATGCGCAAGAAC 1200  
DB 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlnHisIleAlaArgAsn 400  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGCAGGCGGCAAGAGGCGCACAGATGAAG 1260  
DB 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisIleGlnMetLys 420  
QY 1261 GACTGCACCGAGCGCGCAGGCCAACTTCTGTGGGCAAGATCTGCGCGCAGCCACAGGCGCGC 1320  
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
QY 1321 CCGGCAACTTCTGTGAGAACCGGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
DB 441 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454  
QY 1381 CCGCGCGAGAGCTTCGCTTCGAGAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCG 1440  
DB 455 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProIleGlu 474  
QY 1441 CGCGAGCCCTACCGCGAGCGCGCTGACCGCGCTGCGCGCAGCGCGCGCGCGCGCTG 1500  
DB 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
QY 1501 AGCCAG 1506  
DB 491 SerGln 492  
  
RESULT 3  
US-09-991-258-5  
; Sequence 5, Application US/09991258  
; Patent No. US20020141975A1  
; GENERAL INFORMATION:  
; APPLICANT: Olmsted, Robert  
; APPLICANT: Keith, Paula  
; APPLICANT: Dryga, Sergey  
; APPLICANT: Caley, Ian  
; APPLICANT: Maughan, Maureen  
; APPLICANT: Johnston, Robert  
; APPLICANT: Davis, Nancy  
; APPLICANT: Swanstrom, Ronald  
; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE  
; FILE REFERENCE: 01113.0001U3  
; CURRENT APPLICATION NUMBER: US/09/991,258  
; PRIOR FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: 09/902,537  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: 60/216,995  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =  
; OTHER INFORMATION: synthetic construct

US-09-991-258-5

Alignment Scores:

Pred. No.: 2.29e-118 Length: 492  
Score: 2457.00 Matches: 465  
Percent Similarity: 95.22% Conservative: 13  
Best Local Similarity: 92.63% Mismatches: 14  
Query Match: 86.51% Indels: 10  
DB: 9 Gaps: 2

US-09-475-704A-4 (1-1509) x US-09-991-258-5 (1-492)

QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGAGAGCTGACAAGTGGGAGAGATCCGC 60  
Db 1 MetAlaAlaArgAlaSerIleLeuArgGlyGluLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGAGAGACACTACATGCTGAAGCAGCTGTGTGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGGGCTTGGCCCTGAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGCAGCGCGCGCGCTGAGACCGGACAGAGAGCTGCGCAGCTGTACAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCCCTGTACTGCGTGCAGCGCGGACATCGAGGTCCGCGACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGAGACAAGTCCAGCAGAGAGACCGCAGCGCGCAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAAGACCTGCAAGGCGCAGATG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACCAAGGCATCAGCGCGCGCGCATGTAAGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCGCGCGAGGTGATCCCGCATGTTACCGCGCTGAGCGAGGCGCCACCGCGCAGAC 540  
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACAAGATGTTGAACACCGGTGGCGCGCGCAGCGCGCGCATGTCAGATGCTGAAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGCGCTGCACCCCGCTGCAGGCGCGCGCTG 660  
Db 201 ThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCCCGCGCGCGAGTGGCGCGAGCGCGCGCGCGCAGCAGCATGCGCGCGCGCAGCAGCCCTG 720  
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGAGCAGATGCGCTGATGACAGCAACCGCGCGCTGCGCGCGCGCGCAGCATCTCAAG 780  
Db 241 GlnGluGlnIleAlaIleTrpMetThrSerAsnProProIleProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTCTGGGCTGGAACAAGATCGCGGATGTACAGCCCGCTGAGCATCTCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIleLeu 280  
QY 841 GACATCCGCGCAGGGCGCGCAGAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTCAAGAC 900  
Db 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 300  
QY 901 CTGCGCGCGCGAGCAGCGCCAGCAGAGAGCTGAAGAACTGATGACCGAGACCTGTGTG 960

Db ||||| 301 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuVal 320  
QY 961 CAGAAAGCCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGGCCCCGGCGCCACCTG 1020  
Db ||||| 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340  
QY 1021 GAGAGATGATGACCGCTGCGAGGGCGGTGGCGCGCGCGCGCCACAAAGCGCGCTGTG 1080  
Db ||||| 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGAGCCAGGCCCAACAGCGTGAACATCATGATGCAGAGAGCAACTTCAAG 1140  
Db ||||| 361 AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380  
QY 1141 GGGCGCGCGCGCAACGTCAGTGTCTCAACTGCGCGCAAGAGGGCCACATGCGCAAGAAC 1200  
Db ||||| 381 GlyProArgArgIleValLysCysPheAsnLysGlyLysGluGlyHisIleAlaArgAsn 400  
QY 1201 TGGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGCAGGAGGCGCAAGAGGCGCAGATGAAG 1260  
Db ||||| 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420  
QY 1261 GACTGCACCGAGCGCGCAGGCCAACTTCTGCGCAAGATCTGGCCCAAGCCCAAGGGCGCG 1320  
Db ||||| 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
QY 1321 CCGCGCACTTCTCTGCAGAAACCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
Db ||||| 441 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454  
QY 1381 CCGCGCGAGAGCTTCCGCTTGCAGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCG 1440  
Db ||||| 455 ProAlaGluSerPheArgPheGluGluThrProAlaProLysGlnGluProIleGlu 474  
QY 1441 CGCGAGCCCTACCGCGAGCGCGCTGACCGCGCTGCGCAGCCTGTTGGGAGCGCGCGCTG 1500  
Db ||||| 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
QY 1501 AGCCAG 1506  
Db ||||| 491 SerGln 492

RESULT 4

US-10-332-413-8  
; Sequence 8, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swannstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdool  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332, 413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 631  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-10-332-413-8

## Alignment Scores:

Pred. No.:	2.3e-118	Length:	631
Score:	2457.00	Matches:	465
Percent Similarity:	95.22%	Conservative:	13
Best Local Similarity:	92.63%	Mismatches:	14
Query Match:	86.51%	Indels:	10
DB:	16	Gaps:	2

US-09-475-704A-4 (1-1509) x US-10-332-413-8 (1-631)

```
QY 1 ATGGGGCCCGCGCCAGCATCTGCGGGCGAGAGCTGACAAGTGGAGAGATCCGC 60
    |||
Db 31 MetAlaAlaArgAlaSerIleLeuArgGlyGluLeuAspLysTrpGluLysIleArg 50

QY 61 CTGCGCCCGCGCGGCAAGACACTATGCTGAAGCACTGTGTGGGCCAGCCGCGAG 120
    |||
Db 51 LeuArgProGlyGlyblybshIeTYrMetLeuLysHsIleValTrpAlaSerArgL 70

QY 121 CTGAGGGCTTCGCCCTGAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCATC 180
    |||
Db 71 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysblybGlnIle 90

QY 181 ATGAAGCAGCTGCAGCCCGCCCTGCAGACCGGCAACCGAGAGCTGCCAGCCTGTACAC 240
    |||
Db 91 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTYrAsn 110

QY 241 ACCGTGGCCACCTGTACTGCGTGCACCGCGCATCGAGGTCCGCGACCAAGAGGCC 300
    |||
Db 111 ThrValAlaThrLeuTYrCysValHsGluLysIleGluValArgAspThrLysGluAla 130

QY 301 CTGGAACAAGATCGAGAGAGAGCAGAACAAAGTCCCAGCAGAAAGCCCAAGGCCAAGAG 360
    |||
Db 131 LeuAspLysIleGlnGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 150

QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAGAACTGCAGGGCCAGATGCTG 420
    |||
Db 151 AlaAspGlyLysValSerGlnAsnTYrProIleValGlnAsnLeuGlnGlyMetVal 170

QY 421 CACCAAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAGAGCC 480
    |||
Db 171 HsGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGluLysAla 190

QY 481 TTCAAGCCCCGAGGTGATTCCTCATGTTCAACGCTGTGAGCGAGGGCGCCACCCCGCAGAC 540
    |||
Db 191 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 210

QY 541 CTGAACACGATGTTGAACACCGTGGGGGGCCACCGAGCCGCCATGCAAGATGCTGAAGAGC 600
    |||
Db 211 LeuAsnThrMetLeuAsnThrValGlyGlyHsGlnAlaAlaMetGlnMetLeuLysAsp 230

QY 601 ACCATCAACGAGGAGGCGCGAGTGGGACCGCTTGCAACCCCGTGCAAGCGCGGCCCGCTG 660
    |||
Db 231 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHsProValHsAlaGlyProIle 250

QY 661 GCGCCCGGCGAGATGGCGGAGACCCCGCGGGCAGCGACATCGCGGGCGCAACCGACCCCTG 720
    |||
Db 251 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 270

QY 721 CAGAGCAGATCGCCTGATGACCAAGCAACCCCGCTGCGCCGCGGACATCTACAAG 780
    |||
Db 271 GlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTYrLys 290

QY 781 CGGTGATCATCTCGGCGCTGAACAAGATCGTGGATGTACAGCCCGGTGAGCATCTCTG 840
    |||
Db 291 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTYrSerProValSerIleLeu 310

QY 841 GACATCCGCGCAGGCGCCCAAGAGAGCCCTTCGCGACTACGTGAGACCGCTTCTTCAAGAC 900
    |||
Db 311 AspIleArgGlnGlyProLysGluProPheArgAspTYrValAspArgPhePheLysThr 330

QY 901 CTGCGCGCGCAGCAGGCGCAACCCAGACGCTGAAGAACTGATGACCGAGACCCCTGCTGCTG 960
    |||
```

```
Db 331 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeuVal 350

QY 961 CAGAACGCCAACCCCGAAGTGCAGAACCATCTGCGGCTCTCGGCCCCGGCGCACCTTG 1020
    |||
Db 351 GlnAsnAlaAsnProAspCysblybThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 370

QY 1021 GAGGAGATGATGACCGCTGCGCAGGGCGGTGGCGGCCCCGGCCACAGAGGCCCGCTGCTG 1080
    |||
Db 371 GlnGluMetMetThrAlaCysbGlnGlyValGlyGlyProGlyHsLysAlaArgValLeu 390

QY 1081 GCCGAGCGGATGAGCCAGGCCAACAGCGTGAACATCATGATGCAGAAAGCAACTTCAAG 1140
    |||
Db 391 AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetGlnArgSerAsnPheLys 410

QY 1141 GCGCCCGCGCGCAAGCTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCAAAGAC 1200
    |||
Db 411 GlyProArgArgIleValLysCysbPheAsnCysbGlyLysGluGlyHsIleAlaArgAsn 430

QY 1201 TGCGCGCCCCCGCAAGAGGGCTGCTGAACTGCGGCAAGAGGGCCACAGATGAAG 1260
    |||
Db 431 CysArgAlaProArgLysblybGlyCysTrpLysCysGlyLysGluGlyHsIleAlaArgAsn 450

QY 1261 GACTGCACCGAGCGCCAGGCCAACTTCTGTGGCAAGATCTGGCCAGCCACAAAGGGCGC 1320
    |||
Db 451 AspCysThrGluArgGlnAlaAsnPheLeuGlyblybIleTrpProSerHsLysblybArg 470

QY 1321 CCCGCAACTTCTGTGCAGAACCGCAGCGAGCCCGCGCCCCCACCGTGCACCGCCCCC 1380
    |||
Db 471 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 484

QY 1381 CCCGCGAGAGCTTCGCTTCGAGAGAGACCAACCCCGCCCCCAAGAGAGGCCCAAGAG 1440
    |||
Db 485 ProAlaGluSerPheArgPheGlnGluThrThrProAlaProLysGlnGluProIleGlu 504

QY 1441 CGCGAGCCCTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTGCGAGCGGCCCTG 1500
    |||
Db 505 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 520

QY 1501 AGCCAG 1506
    |||
Db 521 SerGln 522
```

## RESULT 5

```
US-10-339-217-109
; Sequence 109, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechler, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; FILE REFERENCE: 111021.143 (ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR APPLICATION NUMBER: 2003-01-09
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 109
; LENGTH: 492
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-109
```

## Alignment Scores:

Pred. No.:	1.71e-117	Length:	492
Score:	2440.00	Matches:	464
Percent Similarity:	94.42%	Conservative:	10
Best Local Similarity:	92.43%	Mismatches:	18
Query Match:	85.92%	Indels:	10

DB: 14 Gaps: 2  
US-09-475-704A-4 (1-1509) x US-10-339-217-109 (1-492)  
QY 1 ATGGGCGCGCGCGCAGCATCTCGCGCGGAGAGAGTGGACAAGTGGAGAGATCCGC 60  
Db 1 MetGlyAlaSerAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCCCGCAG 120  
Db 21 LeuArgProGlyGlyLysLysLysTrpArgLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGGCTTGGCCCTGAACCCCGGCTGTGAGAGCCGCGAGGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaLeuAsnSerGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60  
QY 181 ATGAAGAGCTGCAGCCCGCGCTGCAGACCGGCAACCGAGAGCTGCGCAGCTGTACAC 240  
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyraSn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGCCCGGCATCGAGTCCGCGCAACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrcysValHisAlaGlyIleGluValArgAspThrLysGluAla 100  
QY 301 CTGAGACAGATCGAGAGAGAGAGAGCAAGTCCAGCAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysGlu 120  
QY 361 GCGGAGCGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGTGAGGCGCAGATG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTyrcProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACCAAGCCATCAGCCCGCGCAACCTGTGGTGAAGTGAATCGAGAGAGAGGCC 480  
Db 141 HisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCCCGCGAGGTGATTCCTCATGTTCACCGCTGAGCGAGGGCGCCACCCCGCAG 540  
Db 161 PheSerProGluValIlePheProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGCGCGGCAACGAGCCCGCATGAGATGCTGAAGGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGGAGACCGCTGTGACCCCGTGCAGGCGCGCCGTG 660  
Db 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProVal 220  
QY 661 GCGCGCGCGCGCAGATGCGCGCGCGCGCGCGCAGCAGATCGCGCGCGCGCCAGCAGCCCTG 720  
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGAGACGATCGCTGTGATGAGCAGCAACCCCGCGTGCCTGGGGGAGCATCTACAG 780  
Db 241 GlnGluGlnIleThrTrpMetThrSerAsnProProValProValGlyAspIleTyrls 260  
QY 781 CCGTGGATCATCTGGGGCTGAACAGATCGTGGGATGTACAGCCCGGTGAGCATCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTySerProValSerIleLeu 280  
QY 841 GACATCCGCGAGGGCGCGCAAGAGCGCTTCGCGCACTACGTGAGCGCTTCTTCAAGACC 900  
Db 281 AspIleLysGlnGlyProLysGluProPheArgAspTyrcValAspArgPhePheLysVal 300  
QY 901 CTGCGCGCGCGCAGAGAGCGCAGCAGCAGTGAAGACTGTGATGAGCAGAGCCCTGTGTG 960  
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAAAGCCCAACCCGACTGCAAGACCATCTGTGGCGCTCTGCGCGCGCGCGCAGCCCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProAlaAlaSerLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGGCGGTGGGCGCGCGCGCGCCAGAGCGCGCTGTG 1080

Db 341 GluGluMetMetThrAlaCysGlnGlnGlyValGlyGlyProSerHisLysAlaArgValLeu 360  
QY 1081 GCCGAGGCGATGAGCCAGGCGCAACGCTGAACATCATGATGAGAGAGAGCACTTCAAG 1140  
Db 361 AlaGluAlaMetSerGlnAlaAsnThrThrAsnIleMetMetGlnLysSerAsnPhels 380  
QY 1141 GCGCGCGCGCGCAACGTCAAGTCTTCACTGCGCGCAAGAGGGCCACATCGCGCAAGAAC 1200  
Db 381 GlyProArgArgThrValLysCysPheAsnTyrcGlyLysGlnGlnHisIleAlaLysAsn 400  
QY 1201 TCGCGCGCGCGCGCAAGAGGCTGTGAAGTCCGGCAAGAGAGGGCCACAGATGAAG 1260  
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlnHisIleAlaLysAsn 420  
QY 1261 GACTGCACCGAGCGCCAGGCGCAACTTCTGTGGCAAGATCTGGCCAGCCACAGAGGGCGC 1320  
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
QY 1321 CCGGCGCACTTCTGTGCAAGCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
Db 441 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 454  
QY 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db 455 ProAlaGluSerPheArgPheGlnGluThrThrProAlaProLysGlnGluProLysAsp 474  
QY 1441 CGCGAGCCCTTACCGCGAGCGCGCTGACCGCGCTGTGCGCAGCTGTGCGCAGGGCGCCCTG 1500  
Db 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
QY 1501 AGCCAG 1506  
Db 491 SerGln 492  
RESULT 6  
US-10-339-217-110  
; Sequence 110, Application US/10339217  
; Publication No. US20030198648A1  
; GENERAL INFORMATION:  
; APPLICANT: Buechter, Douglas  
; APPLICANT: Hou, Xiaohong  
; APPLICANT: Marlor, Christopher W.  
; APPLICANT: Rice, William G.  
; APPLICANT: Yang, Wengang  
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of  
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA  
; FILE REFERENCE: 111021.143(ACH-US1)  
; CURRENT APPLICATION NUMBER: US/10/339,217  
; PRIOR FILING DATE: 2003-01-09  
; PRIOR APPLICATION NUMBER: US 60/347,369  
; PRIOR FILING DATE: 2002-01-11  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 110  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: HIV-1  
; US-10-339-217-110  
Alignment Scores:  
Pred. No.: 5.57e-117 Length: 492  
Score: 2430.00 Matches: 458  
Percent Similarity: 94.82% Conservative: 18  
Best Local Similarity: 91.24% Mismatches: 16  
Query Match: 85.56% Indels: 10  
DB: 14 Gaps: 2  
US-09-475-704A-4 (1-1509) x US-10-339-217-110 (1-492)  
QY 1 ATGGGCGCGCGCGCAGCATCTCGCGCGGAGAGAGTGGACAAGTGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspThrTrpGluLysIleArg 20



Db 41 LeuGIuargPheAlaValaAsnProGlyLeuLeuGIuThrAlaGIuGlyCysLysGlnIle 60  
QY 181 ATGAGCAGCTGCAGAGCCCGCCTGCAGACCGGACCGGAGGACTGCGCAGCTGTACAAAC 240  
Db 61 ILeYsGlnLeuGlnProAlaLeuGlnThrGlyThrGIuGlnLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTCTACTGCGTGCACAGCGCGGACATCGAGGTCCGCGACACCAAGAGGCGC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlnIleGluValArgAspThrLysGlnAla 100  
QY 301 CTGGACAAGATCGAGGAGGACAGAAACAAGTCCCGACAGAAAGCCAGCAGGCGCAAGGAG 360  
Db 101 LeuAspArgIleGIuGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaAsnGln 120  
QY 361 GCCGACGGGAAGGTGAGCCAGAACTACCCCATCGTGCAAACTGCAAGGCGCAGATGGTG 420  
Db 121 AlaAspGlyLysValSerGlnAsnLysProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACGAGGCCATCAGCCCCCGCACCCCTGAACGCGCTGGTGAAGGTGATCGAGGAAGGCC 480  
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValIleGIuGlnLysAla 160  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCGCACCCCGCAGGAC 540  
Db 161 PheSerProGIuValIleProMetPheThrAlaLeuSerGIuGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGGGGCGCCACGAGCGCGCCATGACATGCTGAAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGGCGCGCAGTGGGACCGCCTTGCAACCCCGTGCAAGCGCGCGCGCTG 660  
Db 201 ThrIleAsnGIuGlnAlaAlaGIuTyrAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCCCGCGCGCAGATGCGCGGACCCCGCGGCGACGACATGCGCGCGCGCGCACAGACCTG 720  
Db 221 AlaProGIuGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerSerLeu 240  
QY 721 CAGAGCAGATGCGCTGGATGACCAACCCCGCGTGCCTGCGGCGACATCTCAAG 780  
Db 241 GlnGIuGlnIleAlaTyrMetThrGlyAsnProProValProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCTTGAACAAGATCGTGGGATGTACAGCGCGCTTCTTCAAGACC 840  
Db 261 ArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCCAGGCGCGCCCAAGAGAGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900  
Db 281 AspIleLysGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePheLysThr 300  
QY 901 CTGGCGCGCGAGCAGGCGCACCCAGAGACGTGAAGAAGTGGATGACCGAGACCTGCTGGTG 960  
Db 301 LeuArgAlaGIuGlnAlaThrGlnAspValLysAsnTyrMetThrAspThrLeuLeuVal 320  
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCGCTCTCGGCGCGCGCGCACCTCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGCTGGGCGCGCGCGCGCACAGGCGCGCTGCTG 1080  
Db 341 GIuGlnMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGAGCCAGGCGCAACAGCGTGAACATCATGATGACAGAAGCAACTTCAAG 1140  
Db 361 AlaGlnAlaMetSerGlnThrAsnSer--ThrIleLeuMetGlnArgSerAsnPheLys 379  
QY 1141 GGGCCCCCGCGCAACGTCAAGTCTTCAACTGGCGGCAAGGAGGCGCACATGCGCAAGAAC 1200  
Db 380 GlyProLysArgIleValLysCysPheAsnLysGlyLysGlnGlyHisIleAlaLysAsn 399  
QY 1201 TGGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGGCGCGCACGATGAAG 1260  
Db 400 CysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGlnGlyHisGlnMetLys 419

QY 1261 GACTGCACCGAGCGCCAGGCCAATTCTTGGGCAAGATCTTGCGCCAGCCACAGAGGCGCG 1320  
Db 420 AspCysThrGIuArgGlnAlaAsnPheLeuGlyLysIleTyrProSerHisLysGlyArg 439  
QY 1321 CCCGGCAACTTCTTGACGAACCGCAGCGAGCGCGCGCGCCACCGTGCCCAAGCGCGCC 1380  
Db 440 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 453  
QY 1381 CCCGCGGAGAGCTTCGCTTGAGGAGAGACCAACCCCGCGCGCAAGAGAGCCCAAGGAC 1440  
Db 454 ProAlaGlnSerPheArgPheGIuGlnThrThrProAlaLeuGlnGlnGlyProLysAsp 473  
QY 1441 CCGGAGCCCTTACCGCGAGCGCCCTGACCGCGCTGCGCAGCGCTGTGCGCAGCGCGCTG 1500  
Db 474 -----ArgGluProLeuThrSerLeuArgSerLeuPheGlySerAspProLeu 489  
QY 1501 AGCCAG 1506  
Db 490 SerGln 491  
  
RESULT 8  
US-10-339-217-143  
; Sequence 143, Application US/10339217  
; Publication No. US20030198648A1  
; GENERAL INFORMATION:  
; APPLICANT: Buechter, Douglas  
; APPLICANT: Hou, Xiaohong  
; APPLICANT: Marlor, Christopher W.  
; APPLICANT: Rice, William G.  
; APPLICANT: Yang, Wengang  
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of  
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA  
; FILE REFERENCE: 111021.143(ACH-US1)  
; CURRENT APPLICATION NUMBER: US/10/339,217  
; PRIOR FILING DATE: 2003-01-09  
; PRIOR APPLICATION NUMBER: US 60/347,369  
; PRIOR FILING DATE: 2002-01-11  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 143  
; LENGTH: 491  
; TYPE: PRF  
; ORGANISM: HIV-1  
US-10-339-217-143  
  
Alignment Scores:  
Pred. No.: 3.92e-116 Length: 491  
Score: 2413.50 Matches: 456  
Percent Similarity: 95.02% Conservative: 21  
Best Local Similarity: 90.84% Mismatches: 14  
Query Match: 84.98% Indels: 11  
DB: 14 Gaps: 3  
  
US-09-475-704A-4 (1-1509) x US-10-339-217-143 (1-491)  
QY 1 ATGGCGCGCGCGCAGCATCTGCGCGCGGAGAGACTGGACAAGTGGAGAATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTyrGluLysIleArg 20  
QY 61 CTGGCGCGCGCGGCAAGAAGCACTACATGCTGAAGCACTGTGTGGGCCAGCGCGAG 120  
Db 21 LeuArgProGIuGlyLysLysHisTyrMetIleLysHisLeuValTyrAlaSerArgGlu 40  
QY 121 CTGAGGGCTTGCCTGAACCCCGGCTGTGTGAGACCGCGGAGGCTGCAAGCATC 180  
Db 41 LeuGIuargPheAlaLeuAsnProGIuLeuLeuGIuThrSerGIuGlnGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGACGCGCGCGCTTGACACCGGCAACGAGAGCTGGCAGCGCTGTACAAC 240  
Db 61 ILeYsGlnLeuGlnProAlaLeuGlnThrGlyThrGIuGlnLeuArgSerLeuHisAsn 80  
QY 241 ACCGTGGCCACCTCTACTGCGTGCACAGCGCGGACATCGAGGTCCGCGACACCAAGAGGCGC 300

```
Db      81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluIleArgAspThrLysGluAla 100
QY      301 CTGCAACAAGATCGAGAGGAGCAGAAAGTCCAGCAGAAAGACCAGCAGGCCAAGAG 360
Db      101 LeuAspLysIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
QY      361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGAGAACCTGACAGGCCAGATGGTG 420
Db      121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGluMetVal 140
QY      421 CACCAAGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATGAGAGAGAGCC 480
Db      141 HisGlnAlaLeuSerProArgThrLeuAsnAlaTyrValLysValIleGluGluLysAla 160
QY      481 TTCAGCCCCCGAGGTGATCCCATGTTCAACCGCCCTGACGAGGGCGCCACCCCGCAGAC 540
Db      161 PheSerProGluIleIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY      541 CTGAACAAGATGTTGAACACCGTGGGGGGCCACCAAGCCCGCATGACAGATGCTGAAGAC 600
Db      181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200
QY      601 ACCATCAACGAGAGAGCGCCGAGTGGGACCGCCTGCAACCCCGTGACAGCGCGCCGCTG 660
Db      201 ThrIleAsnGluGluAlaAlaGluTyrAspArgLeuHisProAlaGlnAlaGlyProIle 220
QY      661 GCGCGCGCGCAGATGGCGGACCCCGCGGCGCAGCAGCATCGCGCGCGCCACCAAGACCTG 720
Db      221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyIleThrSerThrLeu 240
QY      721 CAGAGCAGATCGCCTGGATGACCAAGACCCCGCGTGGCGCGGACATCTACAG 780
Db      241 GlnGluGlnIleAlaTyrMetThrGlyAsnProProValProValGlyGluIleTyrLys 260
QY      781 CCGTGGATCATCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTG 840
Db      261 ArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280
QY      841 GACATCCGCGCAGGGGCGCCCAAGAGGCGCTTCCGCGACTACGTGAGCCGCTTCTCAAGAC 900
Db      281 AspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 300
QY      901 CTGCGCGCGCAGAGCAGGCCACCCAGAGCGTGAAGACTGGATGACCGAGACCTGCTGGTG 960
Db      301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTyrMetThrAspThrLeuLeuVal 320
QY      961 CAGAACGCCCAACCCCGACTGCAAGACATCTGCGCGCTCTGCGCGCGCGCGCACCTG 1020
Db      321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340
QY      1021 GAGGAGATGATGACCGCCTGCGCAGGGCGTGGCGCGCGCGCGCACAGGCCCGCTGCTG 1080
Db      341 GlnGluMetMetThrAlaCysGlnGlyValGlyIleProSerHisLysValaArgValLeu 360
QY      1081 GCCGAGCGGATGAGCGCAGGCCCAACAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140
Db      361 AlaGlnAlaMetSerGlnThrAsnAsn--SerIleLeuMetGlnArgSerAsnPheLys 379
QY      1141 GCGCGCGCGCGCAAGTCAAGTGTCAACTGCGCGCAAGAGGAGGCCACATGCGCAAGAAC 1200
Db      380 GlyPheLysArgThrValLysCysPheAsnCysGlyLysGluGluHisIleAlaArgAsn 399
QY      1201 TGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGGCGCAAGAGGCGCACAGATGAAG 1260
Db      400 CysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGluGluHisGlnMetLys 419
QY      1261 GACTGCAACGAGCGCGCAGGCCCAACTTCTGGGCAAGATCTGGCCCGCACCAAGGCGCGC 1320
Db      420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTyrProSerHisLysGlyArg 439
QY      1321 CCGCGCAACTTCTGCGAGAACCGCAGGAGCGCGCGCGCGCGCACCGTGGCGCACCGCGCC 1380
```

```
Db      440 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 453
QY      1381 CCGCGCGAGAGCTTCGCTTCGAGAGACCAACCCCGCGCCCAAGCAGAGGCCAAGAC 1440
Db      454 ProAlaGluSerPheArgPheGlnGluThrThrProAlaLeuLysGlnGluGlnLysAsp 473
QY      1441 CGCGAGCCCTACCGCGAGCCCGCTGACCGCCCTGCGCAGCCTGTTGGCAGCGCGCCCTG 1500
Db      474 -----ArgGluProLeuHisSerLeuLysSerLeuPheGlySerAspProLeu 489
QY      1501 AGCCAG 1506
Db      490 SerGln 491
```

## RESULT 9

```
US-10-339-217-105
; Sequence 105, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
; FILE REFERENCE: 111021.143 (ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 105
; LENGTH: 508
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-105
```

```
Alignment Scores:
Pred. No.:      8,99e-115      Length:      508
Score:          2387.00      Matches:      454
Percent Similarity: 92.19%      Conservative: 18
Best Local Similarity: 88.67%      Mismatches: 26
Query Match:      84.05%      Indels:      14
DB:              14      Gaps:      3
```

US-09-475-704A-4 (1-1509) x US-10-339-217-105 (1-508)

```
QY      1 ATGGGCGCGCGCGCAGCATCTCGCGCGCGCAGAGACTGGAGAACTGGAGAACTCCGC 60
Db      1 MetGlyAlaArgAlaSerIleLeuArgGlyThrLysLeuAspAlaTyrGluLysIleArg 20
QY      61 CTGCGCGCGCGCGCAGAGCAAGCACTACATGCTGAAGCACCTGTGTGGCCAGCGCGAG 120
Db      21 LeuArgProGlyGlyLysLysIleTyrMetLeuLysHisLeuValTyrPalaserGlu 40
QY      121 CTGAGGCGCTTCGCCCTGAACCCCGCGCTGTGAAGACCGCGAGCGCTGCAAGCAGATC 180
Db      41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluIleCysLysGlnIle 60
QY      181 ATGAAGCAGCTGAGCGCGCGCGCTGAGACCGCGCAGAGAGCTGCGCAGCTGTACAAC 240
Db      61 MetLysGlnLeuHisProAlaLeuGlnThrGlyThrGluLysSerLeuTyrAsn 80
QY      241 ACCGTGGCAGCCTGTACTGCTGTCAGCGCGCGCATGAGGTCCCGCACCAAGAGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHisGluAsnIleLysValaArgAspThrLysGluAla 100
QY      301 CTGGAACAAGATCGAGAGAGAGCAGAAC-----AAGTCCAGAGCAAGAACCCAGAGGCC 354
Db      101 LeuAspLysIleGluGluGluGlnAsnLysIleLysSerGlnGlnLysThrGlnGlnAla 120
```

```
QY 355 AAGAGCGCGCAGGAGGTGAGCCAGAACTACCCCATCGTGCAAGAACTGCAAGGCGCAG 414
Db 121 LysAlaAlaAspGluValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGln 140
QY 415 ATGTGCAACGAGCCATCAGCCCCCGCAGCCCTGAACGCGCTGGGTGAAGTATCGAGAG 474
Db 141 MetValHisGlnAsnLeuSerProArgThrLeuAsnAlaTrpValValIleGluGln 160
QY 475 AAGGCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCC 534
Db 161 LysAlaPheSerProGluValIlePheMetPheThrAlaLeuSerGluGlyAlaThrPro 180
QY 535 CAGGACTGAAACAGATGTTGAACACCGGTGGCGGCCACGAGCGCCATGCAAGATGCTG 594
Db 181 GlnAspLeuSerThrMetLeuAsnTrpValGlyGlyHisGlnAlaAlaMetGlnMetLeu 200
QY 595 AAGGACCATCAACGAGGAGGCGCGGAGTGGAGCGCGCTGCACCCCGTGACGCGCGC 654
Db 201 LysAspThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGly 220
QY 655 CCGGTGGCCCCCGGCGAGATGCGCGACCCCGCGGCGAGCAGATCGCGCGCGCCACGAG 714
Db 221 ProMetAlaProGluGlnLeuArgGluProArgGlySerAspIleAlaGlyThrThrSer 240
QY 715 ACCCTGCAAGAGCAGATCGCTGTGATGACCAAGCAACCCCGCTGCGCTGGCGCAGATC 774
Db 241 ThrLeuArgGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIle 260
QY 775 TACAAGCGGTGATCATCTCGGCGCTGAACAAGATGCTGGAGTACAGCCCGCTGAGC 834
Db 261 TyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSer 280
QY 835 ATCTTGACATCCGCGCAGGCGCCCAAGAGAGCCCTTCGCGACTACGTGAGCCGCTTCTC 894
Db 281 IleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePhe 300
QY 895 AAGACCTGCGCGCGCAGAGCGCCACCAAGAGCGGTGAAGAACTGATGACCGAGACCTG 954
Db 301 LysAlaLeuArgAlaGluGlnAlaThrGlnAsnValLysAsnTrpMetThrGluThrLeu 320
QY 955 CTGCTGCAGAACCGCCCAACCCGACTGCAGAACCATCTCGCGCTCTCGGCGCGCGCGC 1014
Db 321 LeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyIleGlyAla 340
QY 1015 ACCCTGGAGGAGATGATGACCGCGCTGCAGAGGCGGTGGCGCGCCACAAAGCGCGC 1074
Db 341 ThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArg 360
QY 1075 GTGCTGGCGCGAGCGCATGAGCGCGCCCAACAGCGGTGAACATGATGCAAGAGCAAC 1134
Db 361 ValLeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetMetGlnArgSerAsn 380
QY 1135 TTCAAGGCGCGCGCGCAACGTGATGCTTCACTGCGGCAAGAGGCGCCACATCGCC 1194
Db 381 PheLysSerSerLysArgIleValLysCysSerAsnCysGlyLysGluGlyHisIleAla 400
QY 1195 AAGAACTGCGCGCGCGCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCCACAG 1254
Db 401 ArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 420
QY 1255 ATGAAGAGACTGACCGAGCGCGCCCAACTTCTGGGCAAGATCTGGCCAGCCCAAG 1314
Db 421 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 440
QY 1315 GCGCGCGCGCGCAACTTCTTGCAGAACCGCGAGGAGCGCGCGCGCCACCGCTG----- 1368
Db 441 GlyArgProGluAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSer 460
QY 1369 -----CCCAACCGCGCGCGCGCGCGAGAGCTTCGCTTGCAGAGAGAGC 1410
Db 461 PheArgAsnArgProGluProThrAlaProProAlaGluSerPheArgPheGluGluThr 480
QY 1411 ACCCGCGCGCGCGCAAGAGAGCGCCCAAGAGAGCGCGCTACCGCGAGCGCGCTGACCGCC 1470
```

```
Db 481 ThrProThrProLysGlnGluProLysAsp-----ArgAspProLeuThrSer 496
QY 1471 CTGCGCAGCCTGTTTCGCGAGCGCGCCCTGAGCCAG 1506
Db 497 LeuLysSerLeuPheGlySerAspProSerSerGln 508

RESULT 10
US-10-339-217-104
; Sequence 104, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 104
; LENGTH: 491
; TYPE: PRP
; ORGANISM: HIV-1
US-10-339-217-104

Alignment Scores:
Pred. No.: 1.07e-114 Length: 491
Score: 2385.50 Matches: 453
Percent Similarity: 93.82% Conservative: 18
Best Local Similarity: 90.24% Mismatches: 20
Query Match: 84.00% Indels: 11
DB: Gaps: 3
```

```
US-09-475-704A-4 (1-1509) x US-10-339-217-104 (1-491)
QY 1 ATGGGCGCGCGCGCCAGCATCTCGCGCGCGAGAGCTGGAACAAGTGGAGAAAGATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20
QY 61 CTGCGCGCGCGCGCAAGAACTGATGCTGAAGCACTGTGAGCACTGTGGCGCAGCGCGAG 120
Db 21 LeuArgProGluGlyArgLysHisTyrMetLeuLysHisIleValTrpAlaSerArgLys 40
QY 121 CTGAGGCTTCGCGCTGAACCCCGCGCTGCTGAGACCGCGAGGCTGCAAGCAAGATC 180
Db 41 LeuGluArgPheAlaLeuAsnProGluLeuLeuGluThrSerGlnGlyCysLysGlnIle 60
QY 181 ATGAAGAGCTGACGCGCGCGCTGCAAGACCGGACCGGAGAGTGGCGAGCGCTGTACAAC 240
Db 61 IleLysGlnLeuHisProAlaLeuLysThrGlyThrGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCTGTCGTCAGCGCGCGCATGAGTCCGCGACCAAGAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGlnAsnIleGluValArgAspThrLysGluAla 100
QY 301 CTGACAGAATCGAGGAGGAGCAGAACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 101 LeuAspLysIleGluGluGluGluGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120
QY 361 GCGGAGGCAAGGTGAGCGCAGAACTAACCCCATCGTGCAGAACTGCAAGGCGCAGATGCTG 420
Db 121 AlaAspGluGlyValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140
QY 421 CACCAAGCCATCAGCCCCCGCACCTGAAAGCGCTGGGTGAAGTGAAGTGAAGAGAGAGCC 480
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160
```

```
QY 481 TTCAGCCCCGAGGTGATCCCATGTTGACCGCCCTGAGCGAGGGCGCCACCCCCAGGAC 540
    |||||||
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY 541 CTGAACACGATGTTGAACACCGTGGGGCCACCAAGCCGCCATGACATGCTGAAGAC 600
    |||||||
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuValAsp 200
QY 601 ACCATCAACGAGAGAGCGCCGAGTGGGACCGCCTGACACCCCGTGCAGCGCCGCCGCTG 660
    |||||||
Db 201 ThrIleAsnGluGlnAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProAla 220
QY 661 GCCCCCGCCGAGATGGCGGACCCCGCGGCGACGACATCGCCGCGCGCCACAGCACCTG 720
    |||||||
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240
QY 721 CAGAGCAGATCGCCTGATGACCAACGACACCCCGCTGCGCCGCTGGGGACATCTACAG 780
    |||||||
Db 241 GlnGluGlnIleAlaTrpMetThrGlyAsnProProValProValGlyAspIleTyrLys 260
QY 781 CGGTGATCATCTCGGCTTGAACACAGATCGTGGGATGTACAGCCCGCTGAGCATCTG 840
    |||||||
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280
QY 841 GACATCCGCCGAGGGCCCCAAGAGAGCCCTTCCGCGACTACGTGACCCGCTTCTCAAGAC 900
    |||||||
Db 281 AspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysVal 300
QY 901 CTGGCGCCGCGAGCAGGCCACCCAGACGCTGAAGACTGATGACCCGACACCTGCTGTG 960
    |||||||
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuIle 320
QY 961 CAGAACGCCAACCCCGCACTGCAAGACCATCTCGCGCTCTCGGCCCGCGCCAGCCCTG 1020
    |||||||
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAlaAlaSerLeu 340
QY 1021 GAGAGATGATGACCGCTGCGAGGGCGTGGGGCGCGCGCCACAAAGGCCGCGTGTG 1080
    |||||||
Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360
QY 1081 GCCGAGGCGATGAGCGCCGCAACGCGTGAACATCATGATGCAGAAAGCAACTTCAAG 1140
    |||||||
Db 361 AlaGlnAlaMetSerGlnAlaAsnSer--AsnIleMetMetGlnArgSerAsnPheLys 379
QY 1141 GCGCCCGCGGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATGCGCAAGAAC 1200
    |||||||
Db 380 GlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 399
QY 1201 TGCGCGCGCCCGCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACCATGAAG 1260
    |||||||
Db 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGlnGlyHisGlnMetLys 419
QY 1261 GACTGCACCGGAGCGCCGCAACTTCTGCGCAAGATCTGGCCAGCCACCAAGGGCGCG 1320
    |||||||
Db 420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439
QY 1321 CCGGCAACTTCTGCGAAGACCGCAGCGAGCCCGCGCCCGCCACCGTGCACCGCGCCC 1380
    |||||||
Db 440 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 453
QY 1381 CCGCGCGAGAGCTTCGCTTGAAGAGAACCAACCCCGCCCGCCCAAGAGAGAGCCCAAGAC 1440
    |||||||
Db 454 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluSerLysAsp 473
QY 1441 CGCGAGCCCTACCGCGAGCCCTGACCGCCCTGCGAGCCTGTTGCGAGCGGCGCCCTG 1500
    |||||||
Db 474 -----ArgGluProLeuIleSerLeuLysSerLeuPheGlySerAspProSer 489
QY 1501 AGCCAG 1506
    |||||||
Db 490 SerGln 491
```

```
RESULT 11
US-10-339-217-107
; Sequence 107, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107
; LENGTH: 497
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-107

Alignment Scores:
Pred. No.: 3.5e-114 Length: 497
Score: 2375.50 Matches: 451
Percent Similarity: 93.04% Conservative: 17
Best Local Similarity: 89.66% Mismatches: 28
Query Match: 83.64% Indels: 7
DB: 14 Gaps: 3

US-09-475-704A-4 (1-1509) x US-10-339-217-107 (1-497)
```

```
QY 1 ATGGGCGCCCGCGCGACATCTCTGGCGCGCGAGAACTGGACAAGTGGGAGAAGATCCGC 60
    |||||||
Db 1 MetGlyAlaArgAlaSerIleLeuSerGlyGlyLysLeuAspLysTrpGluArgIleArg 20
QY 61 CTGGCGCCCGCGGCAAGACACTACATGCTGAAGCACTGTGTGGGCCAGCGCGGAG 120
    |||||||
Db 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40
QY 121 CTGGAGGGCTTCGCTTGAACCCCGCGCTGTGAGACCGCGGAGGGCTGCAAGCATC 180
    |||||||
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60
QY 181 ATGAAGCAGCTGAGCGCCCGCTGCAAGACCGGCAACGAGAGCTGCGCAGCTGTAAAC 240
    |||||||
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLeuLeuArgSerLeuPheAsn 80
QY 241 ACCGTGGCCACCTGTACTGCTGCGACGCGCGGCAATGAGGTCCGCGACACCAAGAGGCC 300
    |||||||
Db 81 ThrValAlaThrLeuTyrCysValHisLysGlyIleGluValArgAspThrLysGluAla 100
QY 301 CTGGAACAAGATCAGAGAGAGAGCAACAAGTCCCAAGCAGAAAGCCAGCGCAAGAG 360
    |||||||
Db 101 LeuAspLysIleGluGluGlnGlnAsnLysCysGlnGlnLysAlaGlnGlnAlaLysAla 120
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCTGCAGAACTGCAGAGGGCCAGATGGTG 420
    |||||||
Db 121 AlaAspGluLysValSerGlnAsnTyrProIleValGlnAsnAlaGlnGlyGlnMetVal 140
QY 421 CACCAAGGCATCAGCCCCCGCACCTTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480
    |||||||
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160
QY 481 TTCAGCCCCGAGGTGATCCCAATGTTCAACCGCCTGAGCGAGGGCGCCACCCCGAGGAC 540
    |||||||
Db 161 PheAsnProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY 541 CTGAACACGATGTTGAACACCGTGGCGGCGCACGAGCGCCCATGACATGCTGAAGGAC 600
    |||||||
```

Db 181 LeuAnThrMetLeuAnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGCGCGGAGTGGACCGCTTGACCCCGTGACAGCGCGCCGCTG 660  
Db 201 ThrIleAsnGlnGluAlaAlaGlnTyrAspArgThrHisProValHisAlaGlnProVal 220  
QY 661 GCGCGCGCGGAGATGCGCGGACCGCGCGGAGCGACATGCGCGCGCGGAGCGCGGAG 720  
Db 221 AlaProGlnGlnMetArgGlnProArgGlySerAspIleAlaGlnTyrThrSerThrLeu 240  
QY 721 CAGAGAGATGCGCTGTGATGACCAAGACCCCGCTGCGCGCGGAGCATCTTCAAG 780  
Db 241 GlnGlnGlnIleAlaTyrMetThrSerAsnProProIleProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGTGAGCATCTG 840  
Db 261 ArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCGCGCCAGGCGCGGCAAGAGCGCTTCCGCGGACTACGTGACCGCTTCTCAAG 900  
Db 281 AspIleLysGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePheLysThr 300  
QY 901 CTGCGCGCGGAGAGCGGACCGCAGAGCGTGAAGACTGGATGACCGAGACCTGTGCTG 960  
Db 301 LeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTyrMetThrAspThrLeuVal 320  
QY 961 CAGAGCGCCAAACCGGACTGCAAGACATCTGCGCGCTCTGCGCGCGCGGCGGCACTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCGCGGCGGCAAGAGCGCGCTG 1080  
Db 341 GlnGlnMetMetThrAlaCysGlnGlyValGlyProSerHisLysAlaArgValLeu 360  
QY 1081 GCGGAGCGGATGAGCGGCGGCAAGCGGTGAACATCATGATGACGAAGACCACTTCAAG 1140  
Db 361 AlaGlnAlaMetSerGlnThrAsnSer--AsnIleLeuValGlnArgSerAsnPheLys 379  
QY 1141 GCGCGCGCGGCAAGCTCAAGTCTTCACTGCGCGGCAAGAGGCGGCAATGCGCAAGAC 1200  
Db 380 GlySerAsnArgIleValLysCysPheAsnCysGlyLysValGlnHisIleValArgAsn 399  
QY 1201 TCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGGCGGCAAGAGGCGGCAAGATGAAG 1260  
Db 400 CysArgAlaProArgLysLysGlyCysTyrLysCysGlnGlnGlnHisGlnMetLys 419  
QY 1261 GACTGCACCGAGCGCGGCAAGGCTTCTGCGGCAAGATCTGCGCGGCAAGAGGCGCGC 1320  
Db 420 AspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTyrProSerHisLysGlyArg 439  
QY 1321 CCGCGCAACTTCTTGCAGAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
Db 440 ProGlnAsnPheLeuGlnAsnArgProGlnProThrAlaProProAlaGlnProThrAla 459  
QY 1378 CCGCGCGCGGAGAGCTTCCGCTTGCAGAGAGACCAAGCGCGCGCGGCAAGAGCGG 1437  
Db 460 ProProAlaGlnSerPheArgPheGlnGlnTyrThrProValProLys----- 475  
QY 1438 GACCGCGAGCGCTTACCGCGAGCGCGCTGACCGCGCTGCGAGCTGTGCGAGCGCG 1497  
Db 476 ---ArgGlnLysGlnArgGlnProLeuThrSerLeuLysSerLeuPheGlnAsnAspPro 494  
QY 1498 CTGAGCCAG 1506  
Db 495 SerSerGln 497

RESULT 12  
US-10-339-217-106

; Sequence 106, Application US/10339217  
; Publication No. US20030198648A1  
; GENERAL INFORMATION:  
; APPLICANT: Buechter, Douglas  
; APPLICANT: Hou, Xiaohong

; APPLICANT: Marlor, Christopher W.  
; APPLICANT: Rice, William G.  
; APPLICANT: Yang, Wengang  
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of  
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA  
; FILE REFERENCE: 111021.143(ACH-US1)  
; CURRENT APPLICATION NUMBER: US/10/339,217  
; PRIOR FILING DATE: 2003-01-09  
; PRIOR APPLICATION NUMBER: US 60/347,369  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 106  
; LENGTH: 487  
; TYPE: PRT  
; ORGANISM: HIV-1  
US-10-339-217-106  
  
Alignment Scores:  
Pred. No.: 4.72e-113 Length: 487  
Score: 2353.50 Matches: 448  
Percent Similarity: 94.04% Conservative: 25  
Best Local Similarity: 89.07% Mismatches: 13  
Query Match: 82.87% Indels: 17  
DB: 14 Gaps: 5  
  
US-09-475-704A-4 (1-1509) x US-10-339-217-106 (1-487)  
QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGGCGGAGAGCTGGAACAGTGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuLysGlyLysLysLeuAspThrTyrGlnArgIleArg 20  
QY 61 CTGCGCGCGCGGCAAGAGCACTACATGCTGAAGCACCTGTGTGGCGGCGGCGAG 120  
Db 21 LeuArgProGlnGlyLysLysHisTyrMetLeuLysHisIleValTyrAlaSerArgGln 40  
QY 121 CTGAGGCGCTTCCGCTGAACCGCGCGCTGTGAGACCGCGGAGGCTGCAAGCAGATC 180  
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnTyrAlaGlnGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGCAGCGCGCGCTGCAAGCGGCGGAGAGCTGCGAGCTGTACAAC 240  
Db 61 MetGlnGlnLeuGlnSerAlaLeuGlnTyrGlnGlnGlnLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGCGCACTTGTACTGCGTGCAGCGCGGCGCATGAGGTCCGCGCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisLysGlnIleAspValArgAspThrLysGlnAla 100  
QY 301 CTGACAAGATGAGAGAGAGCAGAAAGTCCCAAGAGAGAGCCAGAGAGCCAGAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaGlnAla 120  
QY 361 GCCGAC---GGCAAGGTGAGCCAGAACTACCCCATCTGTGCAAGACCTGAGGCGCAGATG 417  
Db 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 140  
QY 418 GTGCACAGGCGCATCAGCGCGCGGCGGAGCGCTGAGAGCGCTGAGAGAGAGAGAG 477  
Db 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValIleGlnGlnLys 160  
QY 478 GCCTTACGCGCGGAGTGAATCCCATGTTCACCGCGCTGAGAGAGAGGCGGCGGCGAG 537  
Db 161 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlnAlaThrProGln 180  
QY 538 GACCTGAACAGATGTGAACACCGTGGGCGGCGGAGCGGCGGCGGCGGAGAGAG 597  
Db 181 AspLeuAnThrMetLeuAnThrValGlnGlnHisGlnAlaAlaMetGlnMetLeuLys 200  
QY 598 GACACCATCAACGAGAGGCGCGGAGTGGAGCGGCTGACCCCGTGCAGGCGGCGGCG 657  
Db 201 AspThrIleAsnGlnGlnAlaAlaGlnTyrAspArgLeuHisProValHisAlaGlnPro 220  
QY 658 GTGCGCGCGCGGCGGAGATGCGGAGCGCGCGGCGGAGGAGCATGCGCGGCGGCGGAGCAG 717

```
Db      221 ILeaIaProGIyGlnMetArGluProArGIySerApIleAlaGIyThrThrSerThr 240
QY      718 CTGACAGACAGATCGCCTGGATGACCAACCCCGCTGCCGTGGGCGACATCTAC 777
Db      241 LeuGIInGIInIleAlaTrpMetThrSerAsnProIleProValGIyAspIleTyr 260
QY      778 AAGCGGTGATCATCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 837
Db      261 LysArGTripIleIleuGIyLeuAsnIleValArGMeTyrSerProValSerIle 280
QY      838 CTGACATCCGACGAGCGCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAG 897
Db      281 LeuAspIleLysGIInGIyProLysGIuProPheArGAspTyrValAspArGPhePheLys 300
QY      898 ACCCTGGCGCGGACGAGCGCCACCAAGACGTGAAGAAGCTGATGACCGAGACCCCTGCTG 957
Db      301 ThrLeuArGIaGIuGIInSerSerGIInGIuValLysAsnTrpMetThrAspThrLeuLeu 320
QY      958 GTGACGAACGCCAACCCGCACTGCAAGACCATCTCGCGCTTCCGGCCCGCGCCACC 1017
Db      321 ValGIInAsnAlaAsnProAspCysLysThrIleLeuArGIaLeuGIyProGIyAlaThr 340
QY      1018 CTGACGAGATGATGACCGCCTGCAAGGCGGTGGCGGCCCGCCACAAGCGCCGGTG 1077
Db      341 LeuGIuGIuMetMetThrAlaCysGIInGIyValGIyGIyProGIyHISLysAlaArGIyVal 360
QY      1078 CTGGCGGAGGCGATGAGCCACGAGCCCAAGCGTGAACATCATGATGCAAGAAGCAACTTC 1137
Db      361 LeuAlaGIuAlaMetSerGIInAlaAsnThr--AsnIleMetMetGIInLysSerAsnPhe 379
QY      1138 AAGGCGCCCGCGCGCAACGTCAAGTCTTCACTGCGGCAAGAGGGCCACATCGCCCAAG 1197
Db      380 LysGIyProLysArGIyThrValLysCysPheAsnCyGIyLysGIuGIyHISIleAlaArGIy 399
QY      1198 AACTGCGCGCGCGCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGAGGGCCACACAGATG 1257
Db      400 AsnCyArGIaAlaProArGIyLysGIyCysTrpLysCyGIyLysGIuGIyHISGIInMet 419
QY      1258 AAGGACTGCACCGAGCGCCCAAGCACTTCTGGGCAAGATCTGCGCCCAAGCCACAAGGGC 1317
Db      420 LysAspCysThrGIuArGIInAlaAsnPheLeuGIyLysIleTrpSerTyrLysGIy 439
QY      1318 CGCCCGCGCAACTTCTCTGCAGAAGCCGACGCGCGCGCCCGCCACCGTGGCCCGCCGCC 1377
Db      440 ArGIserGIyAsnPheLeuGIInSerArGIyProGIu-----ProSerAla 453
QY      1378 CCCCCCGCGAGAGCTTCCGCTTCGAGAGACCAACCCCGCGCCCGCCCAAGACGAGCCCAAG 1437
Db      454 ProProAlaGIuSerPheArGPheGIuIn-----ArgGIuProLys 467
QY      1438 GACCGGAGCGCCTTACCGCGAGCCCGCTGACCGCGCCCTGGCGAGCCTGTTCGGCAGCGGCCC 1497
Db      468 AspLysGIuPro-----ProLeuThrSerLeuLysSerLeuPheGIySerAspPro 484
QY      1498 CTGAGCCAG 1506
Db      485 SerSerGIIn 487

RESULT 13
US-10-339-217-108
; Sequence 108, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
; FILE REFERENCE: 11021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
```

```
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 496
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-108

Alignment Scores:
Pred. No.: 9.05e-113 Length: 496
Score: 2348.00 Matches: 447
Percent Similarity: 92.11% Conservative: 20
Best Local Similarity: 88.17% Mismatches: 24
Query Match: 82.68% Indels: 16
DB: 14 Gaps: 4

US-09-475-704A-4 (1-1509) * US-10-339-217-108 (1-496)
QY      1 ATGGGCGCGCGCGCGACGATCTCGCGCGCGAGAGCTGACAGTGGAGAGATCCGC 60
Db      1 MetGIyAlaArGIaSerIleLeuArGIyGIyLysLeuAspLysTrpGIuLysIleArGIy 20
QY      61 CTGGCGCGCGCGCGCAAGACGACTATGCTGAAGCAACCTGTGTGGCGCCAGCGCGAG 120
Db      21 LeuArGIProGIyGIyLysLysLysTyrMetLeuLysHISLeuValTrpAlaSerArGIu 40
QY      121 CTGAGAGGCTTCGCGCTGAAACCCCGCGCTGTGAGACCGCGCGAGCGGTGCAAGCATC 180
Db      41 LeuGIuArGPheAlaLeuAsnSerGIyLeuLeuGIuThrSerAspGIyCysLysGIInIle 60
QY      181 ATGAAGCAGCTGCAGCCCGCGCGCGCGCAAGCCGAGAGAGAGCTTCGCGCAACCTTACAAC 240
Db      61 ILeGIInGIInLeuGIInProAlaLeuLysThrGIyThrGIuGIuLeuArGIserLysTrpAsn 80
QY      241 ACCGTGGCCACCTCTGTACTGCGTGCACGCGCGCATCGAGGTCCCGGACACCAAGAGGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHISAsnAntIleGIuIleArGIAspThrLysGIuAla 100
QY      301 CTGACAGATTCGAGAGAGAGCAACAAGTCCCAAGACGAGAGCCAGAGCCAGCCAG--- 357
Db      101 LeuAspArGIleGIuGIuGIInLysLysCyGIInGIInLysThrGIInGIInLysThr 120
QY      358 -----GAGCGCGAGCGCAAGGTGAGCCCAAGTACCCCATCGTGCAGAACCTG 405
Db      121 GIInGIuValGIuAlaAlaAspGIyLysValSerGIInAsnTyrProIleValGIInAsnLeu 140
QY      406 CAGGCGCAGATGTGCACAGCGCATCAAGCCCGCGCACCCCTGAAGCGCTGGGTGAAGGTG 465
Db      141 GIInGIyGIInMetValHISGIInSerLeuSerProArGIThrLeuAsnAlaTrpValLysVal 160
QY      466 ATCGAGGAGAGGCGCTTCAGCGCGCGAGGTATCCCATGTTCAACCGCGCTGAGGAGGGC 525
Db      161 ILeGIuGIuLysAlaPheSerProGIuIleIlePrometPheThrAlaLeuSerGIuGIy 180
QY      526 GCCACCCCGCAGACCTGAACAGATGTTGAACACCGTGGCGCGCCACAGCGCGCATG 585
Db      181 AlaThrProGIInAspLeuAsnThrMetLeuAsnThrValGIyGIyHISGIInAlaIleMet 200
QY      586 CAGATGCTGAAGACACCATCAAGAGAGCGCGCGAGTGGGAGCGCGCGAGCATGCGCGTG 645
Db      201 GIInMetLeuLysAspThrIleAsnGIuGIuAlaGIuTrpAspArGIySerAspIleArGIy 220
QY      646 CAGCGCGCGCGCGCGCGCGCGCGAGTGGGAGCGCGCGCGCGCGAGCATGCGCGGC 705
Db      221 HISAlaGIyProValAlaProGIyGIInMetArGIuProArGIySerAspIleArGIy 240
QY      706 GCCACGACACCTGCGAGAGAGATGCGCTGATGACGAGCAACCCCGCGTGGCGCGTG 765
Db      241 ThrThrSerAsnLeuGIInGIuGIInIleAsnTrpMetThrAlaAsnProIleProVal 260
```



```
QY 898 ACCCTGCGCGCCGAGCAGGCCACCAGACGTGAAGAACTGGATGACCGAGACCTGCTG 957
    |||||
Db 301 ThrLeuArgAlaGluGlnSerThrGlnGluValLysAsnTrpMetThrAspThrLeu 320
QY 958 GTGCAGAAACGCCAACCCCGACTGCAGACCATCTGCGCGCTCTCGGCCCCGGCGCAAC 1017
    |||||
Db 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340
QY 1018 CTGAGAGAGATGATGACCGCTTGCAGAGCGGTGGCGCGCCCGCCAGAGCGCGCTG 1077
    |||||
Db 341 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 360
QY 1078 CTGCGCCGAGGCGATGAGCCAGCGCCAGACGCGTGAACATCATGATGCAGAGAGCAACTTC 1137
    |||||
Db 361 LeuAlaGluAlaMetSerGlnAlaAsnThr--SerValMetMetGlnLysSerAsnPro 379
QY 1138 AAGGCGCCCGCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGCGCACATCGCCAAG 1197
    |||||
Db 380 LysGlyProArgArgIleValLysCysPheAsnCyseGlyLysGluGlyHisIleAlaArg 399
QY 1198 AACTGCGCGCGCCCGCGCAAGAGGCGTGTGAAGTCCGGCAAGAGGCGCACATCGCAATG 1257
    |||||
Db 400 AsnCybArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleAlaMet 419
QY 1258 AAGGACTGCACCGAGCGCCAGCGCAACTCTCTGGCAAGATCTGGCCAGCCAGCAAGGCG 1317
    |||||
Db 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439
QY 1318 CGCCCGCGCAACTCTCTGCAGAACCGCAGCGAGCCCGCGCCCGCCACCGTGCACCGCC 1377
    |||||
Db 440 ArgProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAla 453
QY 1378 CCCCCCGCGAGAGCTTCGCTTCGAGAGAGACACCCCGCCCGCCAGAGAGCGCCAG 1437
    |||||
Db 454 ProProAlaGluSerPheArgPheGluGluThrThrProGlyGlnLysGlnLysSerLys 473
QY 1438 GACCGCGAGCCCTACCGCGAGCCCTGCAGCGCCCTGCAGCGCTGTTCGAGCGCGCC 1497
    |||||
Db 474 Asp-----ArgGluThrLeuThrSerLeuLysSerLeuPheGlyAsnAspPro 489
QY 1498 CTGAGCCAG 1506
    |||||
Db 490 LeuSerGln 492

RESULT 15
US-10-339-217-18
; Sequence 18, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; FILE REFERENCE: Nucleocapsid 7 Protein to HIV-1 RNA
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 500
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-18

Alignment Scores:
Pred. No.: 3.79e-108 Length: 500
Score: 2258.00 Matches: 432
Percent Similarity: 91.52% Conservative: 32
```

```
Best Local Similarity: 85.21% Mismatches: 27
Query Match: 79.51% Indels: 16
DB: 14 Gaps: 6

US-09-475-704A-4 (1-1509) x US-10-339-217-18 (1-500)
QY 1 ATGGCGCGCCCGCGCAGCATCTTCGCGCGCGAGAGAACTGGACAAGTGGAGAGATCCGC 60
    |||||
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLysLeuAspAlaTrpGluLysIleArg 20
QY 61 CTGCGCGCCCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGCGCCAGCGCGAG 120
    |||||
Db 21 LeuArgProGlyGlyLysLysLysTyArgLysLeuLysHisIleValTrpAlaSerArgGlu 40
QY 121 CTGAGGCGCTTCGCGCTGAACCCCGCGCTGTGAGACCGCGGAGGCGCTGCAAGCAGATC 180
    |||||
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
QY 181 ATGAAGCAGCTGCAGCGCCCGCTGCAGACCGCGCAGAGAGAGTGCAGCGCTGTACAAC 240
    |||||
Db 61 IleGlyGlnLeuGlnProAlaValGlnThrGlySerGluGluLysSerLeuPheAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGCGCGGCAATGCAGGTCCGCGACCAAGAGGCC 300
    |||||
Db 81 ThrValAlaThrLeuTyrcysValHisGluArgIleGluValLysAspThrLysGluAla 100
QY 301 CTGACAAGATCGAGAGAGAGAGCAACAAGTCCAGCAGAGAGACCAAGCGCCAGAGAG 360
    |||||
Db 101 LeuGluLysIleGluGluGluGlnAlaLysSerLysLysAlaGlnGlnAla---Thr 119
QY 361 GCCGAC-----GCCAAGGTGAGCCAGAACTACCCCATCTGTCAGAACTGTCAG 408
    |||||
Db 120 AlaAspThrArgAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGln 139
QY 409 GCGCAGATGTGCAAGCAGCGCATCAGCCCGCGCACTGAACGCTGTGGAAGGTGATC 468
    |||||
Db 140 GlyGlnMetIleHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIle 159
QY 469 GAGGAGAAAGGCTTCAGCCCGCGAGGTGATCCCATGTTACCGCGCTGAGCGAGGCGCC 528
    |||||
Db 160 GluGluLysAlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAla 179
QY 529 ACCCCCGAGACCTGAACACGATGTTGAACACCGTGGCGGCGCACAGCGCGCATGCGAG 588
    |||||
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 199
QY 589 ATGCTGAAGACACCATCAACGAGAGAGCGCGCGAGTGGAGCGCGCTGCACCCCGTGCAG 648
    |||||
Db 200 MetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgLysHisIleProGln 219
QY 649 GCCGCGCCCGTGGCGCGCGCGCGCATGCGCGAGCCCGCGCGAGACATCGCGCGCGCC 708
    |||||
Db 220 AlaGlyProValAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239
QY 709 ACCAGCACCTGCAGAGAGAGATCGCTGTGATGACCAAGACCCCGCGTGGCGGCG 768
    |||||
Db 240 ThrSerThrLeuGlnGluGlnIleGlyTrpMetThrSerAsnProProValProValGly 259
QY 769 GACATCTACAAGCGGTGATCATCTTCGCGCTGAACAAGATCGTGCAGATGTACAGCCCC 828
    |||||
Db 260 GluIleTyrcysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrcSerPro 279
QY 829 GTGAGCATCTTCGACATCCGCCAGAGGCGCCCAAGAGAGCCCTTCGCGGACTACGTGACCGC 888
    |||||
Db 280 ValSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrcValAspArg 299
QY 889 TTCTTCAAGACCTTGGCGCGCGAGCAGCGCCAGCAGACGTGAAGAAGTGAACCGAG 948
    |||||
Db 300 PheTyrcLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsnTrpMetThrGlu 319
QY 949 ACCCTGCTGTGTCAGAACCGCCAGCCCGCAAGTGCAGAACCATCTTCGCGCGTTCGCGCC 1008
    |||||
Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339
```

```

QY 1009 GGGCCACCTGGAGAGATGATGACCGCTGCCAGGGCGGTGGGGCCCCGCCACAAG 1068
Db 340 GlyAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLys 359
QY 1069 GCGCGGTGCTGGCCGAGGCGATGAGCCAGGCCAACAGCGTGAAC-----ATCATGATG 1122
Db 360 AlaArgValLeuAlaGluAlaMetSerGlnAlaThrAsnLeuAsnGlyAlaAlaMetMet 379
QY 1123 CAGAAGACAATTCAAGGGCCCCCGCGCAAGTCAAGTCTTCAACTGGCGCAAGAG 1182
Db 380 GlnArgSerAsnPhelYsglyProArgLysIleValLysCysPheAsnCybglyLysGlu 399
QY 1183 GGGCCACATGCGCAAGAACTGCGCGCGCGCGCGCGCAAGAGGGCTGCTGAAGTGGCGCAAG 1242
Db 400 GlyHisIleAlaArgAsnCybglyAlaProArgLysLysGlyCysTrpLysCybglySer 419
QY 1243 GAGGGCCACGATGAAGGACTGACCGCGCCAGGCCCACTTCTGGGCAAGATCTGG 1302
Db 420 GluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPhelLeuGlyLysIleTrp 439
QY 1303 CCCAGCCACAAGGGCGCGCGCGCGCAACTCTGCAGAACCGCAGCGAGCGCGCGCGCGCC 1362
Db 440 ProSerHisLysGlyArgProGlyAsnPhelLeuGlnAsnArgProGlu----- 455
QY 1363 ACCGTGCCACCGCGCGCGCGCGCGAGAGCTTCCGCTTC---GAGGAGACCACCGCGCC 1419
Db 456 -----ProThrAlaProProAlaGluSerPheGlyPheGlyGluGluIleThrProSer 473
QY 1420 CCCAAGCAGAGGCCCAAGGACCGGAGAGCTTACCGGAGCGCCCTGACCGCCCTGCGCAGC 1479
Db 474 GlnLysGlnGluGlnLysAsnGlnGluPro-----GlnProLeuThrSerLeuArgSer 491
QY 1480 CTGTTGGGCGAGCGCGCGCGCTG 1500
Db 492 LeuPheGlyAsnAspProLeu 498

```

Search completed: March 11, 2005, 16:38:47  
 Job time : 210.272 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:05:25 ; Search time 42.6742 Seconds  
(without alignments)  
5279.323 Million cell updates/sec

Title: US-09-475-704A-4

Perfect score: 2840  
Sequence: 1 atggggcgccgcgcgcagcat.....gcggcccccctgagccagtaa 1509

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_epool\_p/US09475704/runat\_10032005\_140222\_14851/app\_query.fasta\_1.3342  
-DB=Issued\_Patents\_AA -QFMT=fasteran -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=BITS -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09475704 @CGN 1 1 86 @runat\_10032005\_140222\_14851 -NCPU=6 -ICPU=3  
-NO PMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/iaa/PCITUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	86.5	492	4	US-09-991-258-5 Sequence 5, Appl1
2	2246	79.1	500	1	US-08-375-510-1 Sequence 1, Appl1
3	2246	79.1	500	2	US-08-487-657-1 Sequence 1, Appl1
4	2246	79.1	500	4	US-09-309-572-16 Sequence 16, Appl1
5	2246	79.1	500	4	US-09-718-096-16 Sequence 16, Appl1
6	2238	78.8	1350	4	US-09-952-060-35 Sequence 35, Appl1
7	2237	78.8	512	3	US-08-463-210-8 Sequence 8, Appl1
8	2237	78.8	512	3	US-09-124-900-2 Sequence 2, Appl1
9	2237	78.8	512	4	US-08-463-028-8 Sequence 8, Appl1
10	2223	78.3	500	2	US-08-816-155B-45 Sequence 45, Appl1
11	2223	78.3	500	3	US-09-079-587-45 Sequence 45, Appl1
12	2094	73.7	493	4	US-09-952-060-33 Sequence 33, Appl1

13	2018.5	71.1	478	1	US-08-127-499A-11	Sequence 11, Appl1
14	2018.5	71.1	478	1	US-08-483-847-11	Sequence 11, Appl1
15	2011	70.8	437	3	US-08-392-794A-2	Sequence 2, Appl1
16	1949	68.6	512	4	US-09-319-588C-4	Sequence 59, Appl1
17	1748.5	61.6	498	1	US-08-470-202-59	Sequence 59, Appl1
18	1748.5	61.6	498	1	US-08-471-770-59	Sequence 59, Appl1
19	1748.5	61.6	498	2	US-08-468-059-59	Sequence 59, Appl1
20	1748.5	61.6	498	3	US-09-109-916-59	Sequence 59, Appl1
21	1748.5	61.6	498	4	US-09-886-156-59	Sequence 59, Appl1
22	1748.5	61.6	498	4	US-09-886-149-59	Sequence 59, Appl1
23	1748.5	61.6	498	4	US-09-886-150-59	Sequence 59, Appl1
24	1748.5	61.6	498	4	US-09-886-159-59	Sequence 59, Appl1
25	1748.5	61.6	498	4	US-10-326-090-59	Sequence 59, Appl1
26	1726.5	60.8	498	1	US-08-470-202-60	Sequence 60, Appl1
27	1726.5	60.8	498	1	US-08-471-770-60	Sequence 60, Appl1
28	1726.5	60.8	498	2	US-08-468-059-60	Sequence 60, Appl1
29	1726.5	60.8	498	3	US-09-109-916-60	Sequence 60, Appl1
30	1726.5	60.8	498	4	US-09-886-156-60	Sequence 60, Appl1
31	1726.5	60.8	498	4	US-09-886-149-60	Sequence 60, Appl1
32	1726.5	60.8	498	4	US-09-886-150-60	Sequence 60, Appl1
33	1726.5	60.8	498	4	US-09-886-159-60	Sequence 60, Appl1
34	1726.5	60.8	498	4	US-10-326-090-60	Sequence 60, Appl1
35	1671.5	58.9	363	2	US-08-850-049-130	Sequence 130, App
36	1671.5	58.9	363	2	US-08-050-478-130	Sequence 130, App
37	1671.5	58.9	363	3	US-09-414-117-130	Sequence 130, App
38	1671.5	58.9	363	4	US-09-678-437-130	Sequence 130, App
39	1671.5	58.9	363	4	US-09-943-722-130	Sequence 130, App
40	1545.5	54.4	799	1	US-07-648-796A-7	Sequence 7, Appl1
41	1544	54.4	458	1	US-07-648-796A-5	Sequence 5, Appl1
42	1530.5	53.9	337	1	US-07-648-796A-1	Sequence 1, Appl1
43	1514.5	53.3	328	4	US-08-776-188C-77	Sequence 77, Appl1
44	1453.5	51.2	518	4	US-09-206-551-45	Sequence 45, Appl1
45	1410.5	49.7	294	3	US-09-370-368-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1  
US-09-991-258-5  
Sequence 5, Application US/09991258  
Patent No. 6783939  
GENERAL INFORMATION:  
APPLICANT: Olmsted, Robert  
APPLICANT: Kelth, Paula  
APPLICANT: Dryga, Sergey  
APPLICANT: Caley, Ian  
APPLICANT: Maughan, Maureen  
APPLICANT: Johnston, Robert  
APPLICANT: Davis, Nancy  
APPLICANT: Swanson, Ronald  
TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE  
FILE REFERENCE: 0113.0001U3  
CURRENT FILING DATE: 2001-11-16  
PRIOR FILING DATE: 2001-11-16  
PRIOR APPLICATION NUMBER: 09/902,537  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: 60/216,995  
PRIOR FILING DATE: 2000-07-07  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 492  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence; No. 6783939e =  
US-09-991-258-5  
Alignment Scores: 2.32e-141 Length: 492  
Pred. No.: 2457.00 Matches: 465  
Score:

Percent Similarity: 95.22%  
Best Local Similarity: 92.63%  
Query Match: 86.51%  
DB: 4  
Conservative: 13  
Mismatches: 14  
Indels: 10  
Gaps: 2

US-09-475-704A-4 (1-1509) x US-09-991-258-5 (1-492)

QY 1 ATGGGCGCCGCGCCAGCATCTCGCGCGAGAGAGCTGGACAAGTGGAGAGATCCGC 60  
DB 1 MetAlaAlaArgAlaSerIleLeuArgGluGluLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCCGCGCGACAAGACCTACATGCTGAAGACCTGTGTGGCCAGCCGCGAG 120  
DB 21 LeuArgProGluGluLysLysHisTyrMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGGCGCTTCGCCCTGAACCCCGCGCTGTGAGACCGCCGAGGCGCTGCAAGCATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGluLeuGluThrSerGluGluCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGCAGCGCCGCTGCAGACCGGACCGGAGAGCTGCGCAGCTGTACAC 240  
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGluThrGluLysSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGGTGCAGCGCGGATCGAGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGACACAGATCGAGGAGAGACAGAACAGTCCACGAGAGAACCCAGCGCCAGAGAG 360  
DB 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnAlaLysAla 120  
QY 361 GCCGACGCGAAGGTGAGCCAGAACTACCCCATCGTGAGAGACCTGAGGCGCAGATGTG 420  
DB 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGluMetVal 140  
QY 421 CACCAAGCCATCAGCCCGCGCACCCCTGAACGCTGTGAAGTGTATCGAGAGAGGCC 480  
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGGCGCCACCCCGCAGAC 540  
DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGGCGGCCACGAGCCCGCATGAGATGCTGAAGAC 600  
DB 181 LeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGCGCGCGAGTGGGACCGCTGCACCCCGTGCAGGCGCGCCCGTG 660  
DB 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCCCGCGCGCAGATGCGCGCACCCCGCGCGCAGCAGCATCGCGCGCCACCAAGACCTG 720  
DB 221 AlaProGluGlnMetAlaArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGAGCAGATCGCTGTGATGACCAACACCCCGCTGCGCGCGCGCAGACATCTTAAG 780  
DB 241 GlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCGCTGAACAGATCGTGGGATGTACAGCCCGCTGAGCATCTG 840  
DB 261 ArgTrpIleIleLeuGluLysLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGCAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTAAGACC 900  
DB 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 300  
QY 901 CTGCGCGCGCAGAGCGCCACCAAGACGTGAAGAACTGGATGACCGAGACCCCTGTGTG 960  
DB 301 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAACGCCAACCCGACTGCAAGACCATCTGCGCGCTCTGCGCGCCGCGCGCACCTG 1020

DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGluProGluAlaThrLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGGCGGCGCCCGCGCAAGGCGCGTGTG 1080  
DB 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGluHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGATGAGCGCAGGCGCAAGCGTGAACATCATGATGACAGAGGCGCATCGCCAA 1140  
DB 361 AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380  
QY 1141 GCGCGCGCGCGCAACGTCAGTGTCTCAACTGCGCGCAAGAGGCGCACATCGCCAA 1200  
DB 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlnHisIleAlaArgAsn 400  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGGCAAGGAGGCGCACAGATGAG 1260  
DB 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisIleGlnMetLys 420  
QY 1261 GACTGACCGAGCGCGCAGCGCAACTTCTGCGCAAGATCTGGCCAGCGCCAGAGGCGCGC 1320  
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGluLysIleTrpProSerHisLysGlyArg 440  
QY 1321 CCGCGCAACTTCTGCAAGAACCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
DB 441 ProGluAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454  
QY 1381 CCGCGCGAGAGCTTCGCTTGCAGAGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1440  
DB 455 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProIleGlu 474  
QY 1441 CCGGAGCCCTACCGCGAGCGCGCTGACCGCGCTGTGCGCAGCGCGCGCGCGCTG 1500  
DB 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
QY 1501 AGCCAG 1506  
DB 491 SerGln 492  
RESULT 2  
US-08-375-510-1  
; Sequence 1, Application US/08375510  
; Patent No. 5576421  
; GENERAL INFORMATION:  
; APPLICANT: Saito, Atsushi  
; APPLICANT: Sinagawa, Hideo  
; APPLICANT: Nakata, Atsuo  
; TITLE OF INVENTION: HIV ANTIGEN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch and Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/375,510  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/985,949  
; FILING DATE: 04-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svensson, Leonard R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 216-309P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300

TELEFAX: (703) 241-2848  
TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
US-08-375-510-1

Alignment Scores:  
Pred. No.: 1.42e-128 Length: 500  
Score: 2246.00 Matches: 425  
Percent Similarity: 90.16% Conservative: 33  
Best Local Similarity: 83.66% Mismatches: 36  
Query Match: 79.08% Indels: 14  
DB: 1 Gaps: 5

US-09-475-704A-4 (1-1509) x US-08-375-510-1 (1-500)

QY 1 ATGGGCGCCCGCCAGCATCTCGCGCGAGAAAGCTGACAAAGTGGAGAAAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCCCGCGCGCAAGAAAGCACTACATGCTGAAGCACTGTGTGGCCAGCCCGAG 120  
Db 21 LeuArgProGlyGlyLysLysGlnTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGGCTTGGCCCTGAACCCCGCGCTGTGAGAGCCCGAGGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60  
QY 181 ATGAAGCAGCTGACGCGCCCTGCAAGCCGCGCAGAGAGCTGCGCAGCTGTACAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGACGCGCGCATCGAGGTCGCGACCAAGAGGCC 300  
Db 81 ThrIleAlaValLeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAla 100  
QY 301 CTGACAAAGATCGAGAGAGAGCAGAAAGTCCAGCAGAAAGCCAGAGGCCAAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnGlnAsnLysSerLysLysAlaGlnGlnAlaAla 120  
QY 361 GCCGACGCGC-----AAGGTAGCGCAAGAACTACCCCATCGTGCAAACTGACGGC 411  
Db 121 AspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140  
QY 412 CAGATGCTGACACGAGCATCAGCCCGCGCAGCCCTGAACGCTGGTGAAGTGATCAG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValGlu 160  
QY 472 GAGAAAGCCTTACGCGCGAGGTGATCCCATGTTCAACCGCCCTGAGGAGGGCGCCAC 531  
Db 161 GluLysAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnGlyAlaThr 180  
QY 532 CCCAGGACCTGAACAGATGTTGAACCGTGGCGCGCACAGCGCGCATGACATG 591  
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMet 200  
QY 592 CTGAAGACACCATCAAGAGAGGCGCGAGTGGAACGCGCTGACCCCGTGACAGCC 651  
Db 201 LeuLysGlnThrIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHisProValHisAla 220  
QY 652 GGCCCGGTGGCGCCCGCAGATGCGCGACCCCGCGCAGAGCAGATCGCGCGCCAC 711  
Db 221 GlyProIleAlaProGlyGlnMetArgGluProArgLysSerAspIleAlaGlyThrThr 240

QY 712 AGCACCTGACGAGAGATCGCCTGTGATGACCAGCAACCCCGCGTGGCGGAC 771  
Db 241 SerThrLeuGlnGlnGlnIleGlyTyrMetThrHisAsnProIleProValGlyGlu 260  
QY 772 ATCAACAGCGGTGATCATCTGGCGCTGAACAAGATCGTGGATGACAGCCCGTG 831  
Db 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280  
QY 832 AGCATCTGACATCGCCAGGGCCCAAGAGAGCCCTTCGCGACTAGTGACCGCTTC 891  
Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAGACCCCTGCGCGCGAGCAGAGCCCAAGAGAGCCCTTCGCGACTAGTGACCGAGCC 951  
Db 301 TyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTrpMetThrGluThr 320  
QY 952 CTGCTGTGACAAAGCCCAACCCCGCATGCAAGCAATCTGCGGCTCTGCGCGCGC 1011  
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGly 340  
QY 1012 GCCACCTGAGAGATGATGACCGCTGCGCAGCGGTGGCGCGCGCCAGCAAGAGCC 1071  
Db 341 AlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyLysProGlnHisLysAla 360  
QY 1072 CGCGTCTGCGCGCGCATGAGCCAG--GCCAAGAGCTGAACATCATGATGACAAG 1128  
Db 361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLys 380  
QY 1129 AGCACTTCAAGGCGCCCGCGCAAGCTCAAGTCTCAACTGCGCAAGAGAGGCCAC 1188  
Db 381 GlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnGlyLysGlnGlnHis 400  
QY 1189 ATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGCAAGAGGCC 1248  
Db 401 IleAlaLysAsnGlyArgAlaProArgLysLysGlyCysTrpLysGlyLysGlnGly 420  
QY 1249 CACGATGAAGACTGCAACCGAGCGCGCAAGCTTCTGTGGCAAGATCTGCGCCAGC 1308  
Db 421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440  
QY 1309 CACAAGGCGCGCGCGCAACTTCTGCAAGAACCGCAGCGAGCCCGCGCCACCGTG 1368  
Db 441 HisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 454  
QY 1369 CCCACCGCGCGCGCGAGAGCTTCCGCTTC-----GAGAGAACACCCCGCGCCC 1422  
Db 455 ProThrAlaProProGlnGlnSerPheArgPheGlyGlnGlnThrThrProSerGln 474  
QY 1423 AAGCAGAGCCCAAGAGACCGCGAGCCCTACCGCGAGCCCTGACCGCCCTGCGAGCCTG 1482  
Db 475 LysGlnGlnProIleAspLysGlnLeuTyr-----ProLeuAlaSerLeuArgSerLeu 492  
QY 1483 TTCGCGACGCGCGCCCTGAGCCAG 1506  
Db 493 PheGlySerAspProSerSerGln 500

RESULT 3

US-08-487-657-1  
; Sequence 1, Application US/08487657  
; Patent No. 5834267  
; GENERAL INFORMATION:  
; APPLICANT: Saito, Atsushi  
; APPLICANT: Sinagawa, Hideo  
; APPLICANT: Nakata, Atsuo  
; TITLE OF INVENTION: HIV ANTIGEN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch and Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22040-0747

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,657
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,510
; FILING DATE: 18-JAN-1995
; APPLICATION NUMBER: US/07/985,949
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 216-309P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; US-08-487-657-1

```

Alignment Scores:

Pred. No.:	1.42e-128	Length:	500
Score:	2246.00	Matches:	425
Percent Similarity:	90.16%	Conservative:	33
Best Local Similarity:	83.66%	Mismatches:	36
Query Match:	79.08%	Indels:	14
DB:	2	Gaps:	5

US-09-475-704A-4 (1-1509) x US-08-487-657-1 (1-500)

```

QY 1 ATGGGCGCCGCGGCGCATCTGCGCGGAGAGCTGGACAAGTGGAGAGATCCGC 60
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGluLeuAspLysTrpGluLysIleArg 20
QY 61 CTGCGCCCCGCGGCGAAGACATCATGCTGAAGACCTGTGTGGGCCAGCCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysGlnTyrLysLeuYshIleValTrpAlaSerArgGlu 40
QY 121 CTGAGGGCTTGGCCCTGAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGATC 180
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60
QY 181 ATGAAGCAGCTGACGCGCGGCTGAGACCGGACGAGAGCTGCGCAGCTGTACAAC 240
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTGTACTGCGTGCACGCGGCAATCGAGTCCGCGACACCAAGAGGCC 300
Db 81 ThrIleAlaValLeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAla 100
QY 301 CTGACAAGATCGAGAGAGAGACAAGTCCAGACGAGACCAAGAGAGAGAGAGAG 360
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGlnIleAla 120
QY 361 GCCGACGCGC-----AAGGTGACCAAGAACTACCCCATCGTGCAGAACTGCAAGGCGC 411
Db 121 AspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140

```

```

QY 412 CAGATGTGCACCAAGGCCATCAGCCCCCGCACCCCTGAACGCTGGTGAAGTGATCGAG 471
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValValGlu 160
QY 472 GAGAGGCTTACGCCCGGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACC 531
Db 161 GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180
QY 532 CCCAGGACCTGAACACGATGTGAACACCGTGGGCGGCCACGAGCGCCGATGACAGATG 591
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaIleMetGlnMet 200
QY 592 CTGAAGGACACCATCAACGAGAGAGCGCGGAGTGGGACCGGCTGCACCCCGTGCAGGCC 651
Db 201 LeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAla 220
QY 652 GGGCCCGTGGCCCCCGGCGCAGATGCGGACCCCGCGGAGGACATCGCGCGCCACC 711
Db 221 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 240
QY 712 AGCACCTGCAGAGACAGATGCGCTGGATGACCAACAACCCCGCTGCCGTGGCGAC 771
Db 241 SerThrLeuGlnGluGlnIleGlyTrpMetThrHisAsnProProIleProValGlyGlu 260
QY 772 ATCTACAAGCGGTGATCATCTGGGCTGGAACAAGATCGTGGGATGTAAGCCCGCTG 831
Db 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280
QY 832 AGCATCTGCATCCGCGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTC 891
Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300
QY 892 TTCAAGACCTGCGCGCGCGAGCAGCCACCGAGACGTAAGTGAAGTGAACCGAGACC 951
Db 301 TyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGluThr 320
QY 952 CTGCTGTGCAAGACGCAACCCCGCATGCAAGACCATCTGCGCGCTTGGCCCCCGGC 1011
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGly 340
QY 1012 GCCACCTGAGAGATGATGACCGCTGCGCAGGCGGTGGGCGCGCCGCGCAAGAGGCC 1071
Db 341 AlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla 360
QY 1072 CGCGTGTGCGCGGCGGATGAGCCAG---GCCAACAGCGTGAACATCATGATGCAGAG 1128
Db 361 ArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLys 380
QY 1129 AGCAACTTCAAGGCGCCCGGCGCAACGTCAGTCAAGTCACTGCGGCAAGAGGCGCA 1188
Db 381 GlnAsnPheArgAsnGlnIleArgLysThrValLysCysPheAsnCysGlyLysGluGln 400
QY 1189 ATGCCAAGAACTGCGCGCGCCCGCGCAAGAGGCTGTGAAGTGGCAAGAGAGGCGC 1248
Db 401 IleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGly 420
QY 1249 CACCAGATGAAGACTGCAACGAGCGCAAGCCAACTTCTGGCAAGATCTGGCCCAAC 1308
Db 421 HisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440
QY 1309 CACAAGGCGCGCCCGGCAACTTCTGCAAGACCGCAGCGAGCCCGCGCCCAACGCTG 1368
Db 441 HisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 454
QY 1369 CCCACCGCGCGCGCGCGGAGAGCTTCCGCTTC-----GAGAGACCAACCCCGCGCCC 1422
Db 455 ProThrAlaProProGluGluSerPheArgPheGlyGluGluThrThrProSerGln 474
QY 1423 AAGCAGAGCCCAAGAGCGGAGCGCTTACCGGAGCGCCCTGACCGCCCTGCGCAGCCTG 1482
Db 475 LysGlnGluProIleAspLysGluLeuTyr-----ProLeuAlaSerLeuArgSerLeu 492
QY 1483 TTGCGCAGCGCGCCCTGAGCCAG 1506

```

Db 493 PheGlySerAspProSerSerGln 500

## RESULT 4

US-09-309-572-16  
; Sequence 16, Application US/09309572

; Patent No. 6440730

; GENERAL INFORMATION:

; APPLICANT: Heinrich-Pette-Institut

; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV

; FILE REFERENCE: P50489

; CURRENT APPLICATION NUMBER: US/09/309,572

; CURRENT FILING DATE: 1999-05-11

; EARLIER APPLICATION NUMBER: DE 198 56 463

; EARLIER FILING DATE: 1998-11-26

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 16

; LENGTH: 500

; TYPE: PRT

; ORGANISM: Human immunodeficiency virus type 1

; FEATURE:

; OTHER INFORMATION: gag polyprotein

US-09-309-572-16

## Alignment Scores:

Pred. No.:	1.42e-128	Length:	500
Score:	2246.00	Matches:	425
Percent Similarity:	90.16%	Conservative:	33
Best Local Similarity:	83.66%	Mismatches:	36
Query Match:	79.08%	Indels:	14
DB:	4	Gaps:	5

US-09-475-704a-4 (1-1509) x US-09-309-572-16 (1-500)

QY 1 ATGGGGCGCCGCGCCGACATCTGCGCGCGAGAAAGCTGGACAAGTGGGAGAAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGCAAGACACTACTGCTGAAGACACTGTGTGGCCAGCCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysGlnTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGGGGCTTGGCCCTGAACCCCGGCTGTGTGAGACCGCGAGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGlnTrpSerGluGlyCysArgGlnIle 60  
QY 181 ATGAAGCAGCTGACGCGCGCCCTGACAGCCGCGACCGAGAGCTGCGAGCCTGTACAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnTrpGlySerGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCTGCTGACGCGCGGACCTGAGGTCCGCGACCAAGAGGCG 300  
Db 81 ThrIleAlaValLeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAla 100  
QY 301 CTGAACAAGATCGAGAGAGACAGAAAGTCCACAGAGAACCCAGCAGGCGCAAGAG 360  
Db 101 LeuAspLysIleGluGluGlnGlnAsnLysSerLysLysLysAlaGlnGlnAlaAla 120  
QY 361 GCCGACGCGC-----AAGGTAGCGAGAACTACCCCATCTGTCAGAACTGACGCGC 411  
Db 121 AspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140  
QY 412 CAGATGGTGACACGAGGCGCATCAGCCCGCGACCCCTGAACGCTGGGTGAAGGTGACG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleValLysValValGlu 160  
QY 472 GAGAGGCGCTTACGCGCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCGCAC 531  
Db 161 GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180  
QY 532 CCCAGGACCTGAACAGATGTTGAACACCGTGGCGCGCACCGCGCGCATGACGATG 591  
|||||

Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMet 200  
QY 592 CTGAAGACACCATCAACGAGAGGCGCGCGAGTGGACCGCTGCACCCCGTGCAGGCG 651  
Db 201 LeuLysGluThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisProValHisAla 220  
QY 652 GGGCGCGTGGCGCGCGCGAGATGCGCGGACCCCGCGCGAGCAGACATCGCGCGCGCAC 711  
Db 221 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 240  
QY 712 AGCACCTGACAGAGATCGCTGGATGACCAAGCAACCCCGCGTGGCGCGCGAG 771  
Db 241 SerThrLeuGlnGlnIleGlyTyrMetThrHisAsnProProIleProValGlyGlu 260  
QY 772 ATCTACAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGATGTACAGCCCGTG 831  
Db 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280  
QY 832 AGCATCTGACATCCGCGCGAGCGCGCAAGAGAGCCCTTCGCGACTACGTGACCGCTTC 891  
Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAAGACCTGCGCGCGCGAGCGCGCAACCCGAGCTGCAAGACCATCTGCGCGCGCGC 951  
Db 301 TyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGluThr 320  
QY 952 CTGCTGGTGACAGAACCGCAACCCCGAGCTGCAAGACCATCTGCGCGCGCGCGCGC 1011  
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGly 340  
QY 1012 GCCACCTTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCGCGCGCACAAAGCC 1071  
Db 341 AlaThrLeuGluGlnMetMetThrAlaCysGlnGlyValGlyGlyProGluHisLysAla 360  
QY 1072 CGCGTCTGCGCGCGCGATGAGCGAG--GCCAAGCGTGAACATCATGATGACAGAG 1128  
Db 361 ArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLys 380  
QY 1129 AGCACTTCAAGGCGCGCGCGCGCAAGTCAAGTCTCACTGCGCGCAAGAGGCGCAC 1188  
Db 381 GlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHis 400  
QY 1189 ATCGCCAAGACTGCGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTCCGCGCAAGAGGCG 1248  
Db 401 IleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGly 420  
QY 1249 CACCAATGAAGACTGACACCGAGCGCGCGCAACTTCTGCGGCAAGATCTGCGCGCAC 1308  
Db 421 HisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440  
QY 1309 CACAAGGCGCGCGCGCGCACTTCTGACAGAACCGCGAGCGCGCGCGCGCGCGCGTG 1368  
Db 441 HisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 454  
QY 1369 CCCACCGCGCGCGCGCGCGAGAGCTTCCGCTT-----GAGAGACCAACCCCGCGCGC 1422  
Db 455 ProThrAlaProProGluGlnSerPheArgPheGlyGluGluThrThrThrProSerGln 474  
QY 1423 AAGCAGAGCCCAAGAGACCGCGAGCCCTACCGCGAGCCCGCTGACCGCGCGCGCGCTG 1482  
Db 475 LysGlnGluProIleAspLysGluLeuTyr-----ProLeuAlaSerLeuArgSerLeu 492  
QY 1483 TTCGGCAGCGCGCGCGCTGAGCCAG 1506  
Db 493 PheGlySerAspProSerSerGln 500  
|||||

## RESULT 5

US-09-718-096-16

; Sequence 16, Application US/09718096

; Patent No. 6589763

; GENERAL INFORMATION:

; APPLICANT: Von Laer, Melike-Dorothee

; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV

FILE REFERENCE: 35-195  
CURRENT APPLICATION NUMBER: US/09/718,096  
CURRENT FILING DATE: 2000-11-22  
PRIOR APPLICATION NUMBER: DE 19856463.5  
PRIOR FILING DATE: 1998-11-26  
PRIOR APPLICATION NUMBER: EP 99250415.9  
PRIOR FILING DATE: 1999-11-25  
PRIOR APPLICATION NUMBER: US 09/309,572  
PRIOR FILING DATE: 1999-05-11  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 16  
LENGTH: 500  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus type 1  
FEATURE:  
OTHER INFORMATION: gag polypeptide  
US-09-718-096-16

Alignment Scores:  
Pred. No.: 1.42e-128 Length: 500  
Score: 2246.00 Matches: 425  
Percent Similarity: 90.16% Conservative: 33  
Best Local Similarity: 83.66% Mismatches: 36  
Query Match: 79.08% Indels: 14  
DB: 4 Gaps: 5

US-09-475-704A-4 (1-1509) x US-09-718-096-16 (1-500)

QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGGAGAGCTGACAGTGGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
Db 21 LeuArgProGlyGlyLysLysGlnTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGGGGCTTGGCCCTGAACCCCGGCTGCTGAGAGAGAGAGAGAGAGAGAG 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
QY 181 ATGAGGAGCTGACAGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCGCGCGCGCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
Db 81 ThrIleAlaValLeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAla 100  
QY 301 CTGGAAGAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysAlaGlnGlnAlaIleAla 120  
QY 361 GCCGAGCGC-----AAGGTGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 411  
Db 121 AspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140  
QY 412 CAGATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValValGlu 160  
QY 472 GAGAGAGCGCTTACGCGCGCGAGGTGATCCCATGTTACCGCGCTGAGAGAGAGAG 531  
Db 161 GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180  
QY 532 CCCAGGAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591  
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaIleMetGlnMet 200  
QY 592 CTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651  
Db 201 LeuLysGluThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisAla 220

QY 652 GGCCCGGTGGCCCGCGCGAGATGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 711  
Db 221 GlyProIleAlaProGlyGlnMetArgLysProArgLysSerAspIleAlaGlyThrThr 240  
QY 712 AGCAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771  
Db 241 SerThrLeuGlnGlnGlnIleGlyTyrMetThrHisAsnProProIleProValGlyGlu 260  
QY 772 ATCTAAGAGCGGTGATCATCTGCGCGCTGAACAGAGAGAGAGAGAGAGAGAG 831  
Db 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280  
QY 832 AGCATCTGACATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAAGAGCTTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951  
Db 301 TyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGluThr 320  
QY 952 CTGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGly 340  
QY 1012 GCCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071  
Db 341 AlaThrLeuGluGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla 360  
QY 1072 CGCGTCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128  
Db 361 ArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLys 380  
QY 1129 AGCACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188  
Db 381 GlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnLysGlyLysGluGlyHis 400  
QY 1189 ATGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248  
Db 401 IleAlaLysAsnLysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGly 420  
QY 1249 CACGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308  
Db 421 HisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440  
QY 1309 CACAAGGCGCGCGCGCAACTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368  
Db 441 HisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 454  
QY 1369 CCCAGCGCGCGCGCGAGAGCTTCCGCTTC-----GAGAGAGAGAGAGAGAGAG 1422  
Db 455 ProThrAlaProProGluGluSerPheArgPheGlyGluGluThrThrThrProSerGln 474  
QY 1423 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
Db 475 LysGlnGluProIleAspLysGluLeuTyr-----ProLeuAlaSerLeuArgSerLeu 492  
QY 1483 TTGGCAGCGCGCGCTGAGCCAG 1506  
Db 493 PheGlySerAspProSerSerGln 500

RESULT 6  
US-09-952-060-35  
Sequence 35, Application US/09952060  
Patent No. 6733993  
GENERAL INFORMATION:  
APPLICANT: Emini, Emilio A.  
APPLICANT: Youil, Rima  
APPLICANT: Bett, Andrew J.  
APPLICANT: Chen, Ling  
APPLICANT: Kaslow, David C.  
APPLICANT: Shiver, John W.  
APPLICANT: Toner, Timothy J.  
APPLICANT: Casimiro, Danilo R.

```
/ TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
/ TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV-1 GAG, POL, NEF AND
/ FILE REFERENCE: 20747Y
/ CURRENT APPLICATION NUMBER: US/09/952,060
/ CURRENT FILING DATE: 2001-09-14
/ PRIOR APPLICATION NUMBER: PCT/US01/28861
/ PRIOR FILING DATE: 2001-09-14
/ PRIOR APPLICATION NUMBER: 60/317,814
/ PRIOR FILING DATE: 2001-09-07
/ PRIOR APPLICATION NUMBER: 60/279,056
/ PRIOR FILING DATE: 2001-03-27
/ PRIOR APPLICATION NUMBER: 60/233,180
/ PRIOR FILING DATE: 2000-09-15
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 35
/ LENGTH: 1350
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Codon optimized gag-1A pol fusion
US-09-952-060-35

Alignment Scores:
Pred. No.: 5.01e-128 Length: 1350
Score: 2238.00 Matches: 423
Percent Similarity: 90.35% Conservative: 36
Best Local Similarity: 83.27% Mismatches: 35
Query Match: 78.80% Indels: 14
DB: 4 Gaps: 5

US-09-475-704A-4 (1-1509) x US-09-952-060-35 (1-1350)

QY 1 ATGGGCGCCGCGCAGCATCTCGCGCGAGAGCTGGAACAAGTGGGAGAAGATCCGC 60
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyIleuAspLysTrpGluLysIleArg 20
QY 61 CTGCGCGCCGCGCAGCAAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 121 CTGAGGCGCTTCCGCTGAACCCCGGCTGCTGAGACCGCGGAGGCTGCAAGCAGATC 180
Db 41 LeuGluArgPheAlaValAsnProGlyLeuGluThrSerGluGlyCysArgGlnIle 60
QY 181 ATGAAGCAGCTGACGCGCCCTGCAGACCGGACCGAGAGCTGCGACCTGTACAAC 240
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGCGCACCTGTACTGCTGCTGCAACCGCGCATCGAGTCCGCGACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGlnLysIleAspValLysAspThrLysGlnAla 100
QY 301 CTGACAGATCGAGAGGAGAGCAAGCAACTCCAGCAGAGAACCCAGAGGCCAAGAG 360
Db 101 LeuGluLysIleGluGluGlnAsnLysSerLysLysLysLysLysLysLysLysLys 120
QY 361 GCCGACGCGC-----AAGGTAGCCGCAACTACCCCATCTGTGACAGAACTGACGGGC 411
Db 121 GlyThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140
QY 412 CAGATGTGCACACGAGCCATCAGCCCGGCAACCTGAACGCTGGTGAAGGTGATCGAG 471
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleTrpValLysValValGln 160
QY 472 GAGAGGCGCTTCAAGCGCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCCACC 531
Db 161 GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180
QY 532 CCCGAGACCTGAACACGATGTTGAACACCGGTGGCGGCGCACGAGGCCCATGACAGTG 591
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMet 200
```

```
QY 592 CTGAAGCACCATCAACGAGAGGCGCGGAGTGGAACCGCTGCACCCCGTGACAGCC 651
Db 201 LeuLysGluThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisAla 220
QY 652 GCGCCCGTGGCCCCCGCGCAGATGCGGACCCCGCGGAGCGACATCGCGCGCCACC 711
Db 221 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 240
QY 712 AGCACCTGCAGAGAGCATCGCTGTGACCAAGCAACCCCGCTGCGCGGTGGCGAC 771
Db 241 SerThrLeuGlnGluGlnIleGlyTyrMetThrAsnAsnProProIleProValGlyGlu 260
QY 772 ATCTACAGCGGTGATCATCTCGGCTGAACAAGATCGTGGCATGTACAGCCCGTG 831
Db 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280
QY 832 AGCATCTGCAGACATCCGCGAGGCGCCCAAGAGAGCCCTCCGCACTACGTGACCGCTTC 891
Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300
QY 892 TTCAAGACCTTGGCGCGCGAGCAGAGCCCAAGAGAGCTGAAAGACTGTGACCGAGACC 951
Db 301 TyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGluThr 320
QY 952 CTGCTGTGTCAGAACCGCCCAACCCCGACTGCAGACCATCTGCGCGCTTCGCGCCGCGC 1011
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340
QY 1012 GCCACCTTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGGCGCCGCGCACAGGCC 1071
Db 341 AlaThrLeuGluGluMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla 360
QY 1072 CGCGTGTGCGCGAGGCGATGAGCCAG--GCCAACAGCTGAACATCATGTGACAGAG 1128
Db 361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArg 380
QY 1129 AGCAACTTCAAGGCGCGCGCGCAAGCTCAAGTGTCACTGCGGCAAGAGAGGCCAC 1188
Db 381 GlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysValGlyHis 400
QY 1189 ATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGGCGCAAGAGGCGC 1248
Db 401 IleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGly 420
QY 1249 CACCAATGAAGGATGTCACCGAGCGCGCAACTTCTGCGCAAGATCTGGCCAGC 1308
Db 421 HisGlnMetLysAspCysAsnGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440
QY 1309 CACAAGGCGCGCGCGCAACTTCTGCGCAAGACCGCAGCGAGCCCGCGCCACCGTG 1368
Db 441 HisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 454
QY 1369 CCCACCGCGCGCGCGCGAGAGCTTCCGCTTC-----GAGAGACCAACCCCGCGCCC 1422
Db 455 ProThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGln 474
QY 1423 AAGCAGAGCCCAAGAGACCGGAGCCCTACCGCGAGCCCTGACCGCGCTGCGAGCGTG 1482
Db 475 LysGlnGluProIleAspLysGluLeuTyr-----ProLeuAlaSerLeuArgSerLeu 492
QY 1483 TTGCGCAGCGCGCGCTGAGCCAG 1506
Db 493 PheGlyAsnAspProSerSerGln 500

RESULT 7
US-08-463-210-8
/ Sequence 8, Application US/08463210
/ Patent No. 6001977
/ GENERAL INFORMATION:
/ APPLICANT: CHANG, Nancy T.
/ APPLICANT: GALLO, Robert C.
/ APPLICANT: WONG-STAL, Flossie
```

TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan & Finnegan, L.L.P.  
STREET: 345 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10154-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,210  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 436  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 06/693,866  
FILING DATE: 23-JAN-1985  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 06/659,339  
FILING DATE: 10-OCT-1984  
ATTORNEY/AGENT INFORMATION:  
NAME: Serunian, Leslie A.  
REGISTRATION NUMBER: 35,353  
REFERENCE/DOCKET NUMBER: 2026-4193US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 512 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: HTLV-III  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..512  
OTHER INFORMATION: /note= "gag protein of HTLV-III"  
US-08-463-210-8

Alignment Scores:  
Pred. No.: 4.99e-128 Length: 512  
Score: 2237.00 Matches: 422  
Percent Similarity: 89.69% Conservative: 39  
Best Local Similarity: 82.10% Mismatches: 39  
Query Match: 78.77% Indels: 14  
DB: 3 Gaps: 5

US-09-475-704A-4 (1-1509) x US-08-463-210-8 (1-512)

QY 1 ATGGGCGCCGCGGCATCTGCGCGCGGAGAGCTGGACAAGTGGGAGAAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValIleuSerGlyGluIleuAspArgTrpGluIleArg 20  
QY 61 CTGGCCCCCGCGCAAGAAGACATGCTGAAGCACTGTGTGGGCCAGCCCGGAG 120  
Db 21 LeuArgProGlyGlyIleuSerGlyTyrIleuSerHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAGGGCTTGGCCCTGAACCCCGGCTGTGAGACCGCGAGGGCTGCAAGACATC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyIleuGluThrSerGluGlyCysArgGlnIle 60  
QY 181 ATGAAGCAGCTGACGCGCCCTGACAGACCGGCACGAGAGAGCTGGCAGCCTGACAAAC 240  
Db 61 LeuGlyGlnIleuGlnProSerIleuGlnThrGlySerGluGluIleuArgSerLeuTyrAsn 80

QY 241 ACCGTGGCCACCTGTACTGCGTGACCGCGGCATTCGAGTCCGCGACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleuAspThrIleGluVal 100  
QY 301 CTGACAAGATCGAGAGAGACAGAACAAAGTCCAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspIleuSerGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120  
QY 361 GCCGACGCGC-----AAGGTAGCCAGAACTACCCCATCTGTCAGAACCTGACGCGC 411  
Db 121 AspThrGlnHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly 140  
QY 412 CAGATGCTGACAGAGCCCATCAGCCCGCAGCCCTGAACGCTGGGTGAAGTGAATCGAG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrIleuAsnAlaTrpValIleValGlu 160  
QY 472 GAGAGGCTTCAAGCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGCGCGCACCC 531  
Db 161 GluIleuValPheSerProGluValIleProMetPheSerAlaIleuSerGluGlyAlaThr 180  
QY 532 CCCAGAGACCTGAACAGATGTTGAACACCGTGGCGCGGCAGCGCCGCGCATGAGATG 591  
Db 181 ProGlnAspLeuAsnThrMetIleuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMet 200  
QY 592 CTGAAGGACACCAATCAACGAGAGAGCGCGCATGAGAGCGCGCTGACCCCGTGAGGCC 651  
Db 201 LeuIleuSerGluThrIleAsnGluGluAlaIleGluTrpAspArgValHisProValHisAla 220  
QY 652 GCGCCCGTGGCCCGCGGCAGATGCGCAGACCCCGCGCGCAGCGCATGCGCGCGCACCC 711  
Db 221 GlyProIleAlaProGluGluGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 240  
QY 712 AGCAACCTGACAGAGCAGATCGCTGATGACCAACCCCGCTGCGCGCGCGAG 771  
Db 241 SerThrLeuGlnGluGlnIleGlyTyrMetThrAsnAsnProIleProValGlyGlu 260  
QY 772 ATCTACAAGCGGTGATCATCTGGGCTGAAACAAGATCGTGGAGATGACAGCCCGTG 831  
Db 261 IleTyrIleAspArgTrpIleIleuGluIleuAsnIleValAlaGlnMetTyrSerProThr 280  
QY 832 AGCATCTGACATCCGCGAGGCGCCCAAGAGCGCTTCCGCGCATGAGCGCGCTTC 891  
Db 281 SerIleuAspIleArgGlnGlyProIleuGluProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAAGACCTCGCGCGCGCAGAGAGCGCCACCGCAAGCGTGAAGAACTGATGACGAGACC 951  
Db 301 TyrIleThrLeuArgAlaGluGlnAlaSerGlnGluValIleAsnTrpMetThrGluThr 320  
QY 952 CTGCTGTGTGACAAAGCCCAAGCGCTGACCAACCATCTCGCGCGCTCTCGCGCGCGC 1011  
Db 321 LeuIleuValGlnAsnAlaAsnProAspCysIleuThrIleuIleuValAlaLeuGluProAla 340  
QY 1012 GCCACCTGAGAGATGATGACCGCTGCGAGCGCGTGGCGCGCGCGCGCGCAAGGCC 1071  
Db 341 AlaThrLeuGluGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisIleValAla 360  
QY 1072 CCGGTGCTGGCGGAGCGGATGAGCCAG--GCCAAGCGGTGAACATCATGATGACAGAAG 1128  
Db 361 ArgValIleuAlaGluAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGlnArg 380  
QY 1129 AGCAACTTCAAGGCGCCCGCGCAACGTCAGTCTTCACTGCGGCAAGAGAGCGCCAC 1188  
Db 381 GlyAsnPheArgAsnGlnArgGlyMetValIleCysPheAsnIleCysGlyIleGluGlnHis 400  
QY 1189 ATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGCTGTGAGAGTGCAGAGAGAGGCGC 1248  
Db 401 ThrAlaArgAsnIleCysArgAlaProArgIleuGlySerTrpIleuGlySerGlyIleGlu 420  
QY 1249 CACCAAGATGAAGACTGACAGGAGCGCGCAAGCTTCTGGGCAAGATCTGGCCAGC 1308  
Db 421 HisGlnMetIleuAspCysThrGluArgGlnAlaAsnIleuGluGlyIleTyrProSer 440  
QY 1309 CACAAGGCGCGCGCGCAACTTCTGACAGAACCGGAGGAGCGCGCGCGCGCCCAACCGTG 1368

```

Db      441  TyrIysGlyArgProGlyAsnPhenLeuGlnSerArgProGluProThrAlaProProPhe 460
      :::||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||||
Qy      1369 -----CCGACCGCCCCCCCCCGCGAGAGCTTCGC-----TCGAG 1404
      |||||
Db      461  LeuGlnSerArgProGluProThrAlaProProGluGlnSerPheArgSerGlyValGlu 480
      |||||
Qy      1405  GAGACCACCCCCGCCCCAAGCAGAGACCCCAAGACCAGCGCCCTACCGCGAGCCCTG 1464
      |||||
Db      481  ThrThrThrProProGlnLysGlnGluProIleAspLysGluLeuTyr-----ProLeu 498
      |||||
Qy      1465  ACCGCCCTGCGCAGCCTGTTCGCGCAGCGGCCCCCTGAGCCAG 1506
      |||||
Db      499  ThrSerLeuArgSerLeuPheGlyAsnAspProSerSerGln 512

```

RESULT 8  
US-09-124-900-2  
; Sequence 2, Application US/09124900

```

; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andrea
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: PREDL, Renate
; APPLICANT: SCHMATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112P
; CURRENT APPLICATION NUMBER: US/09/124,900
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: PCT/EP95/01481
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-124-900-2

```

Alignment Scores:	
Pred. No.:	4.99e-128
Score:	2237.00
Percent Similarity:	89.69%
Best Local Similarity:	82.10%
Query Match:	78.77%
DB:	
Gaps:	3
Matches:	512
Conservative:	422
Mismatches:	39
Indels:	14
Gaps:	5

US-09-475-704A-4 (1-1509) x US-09-124-900-2 (1-512)

```

QY      1 ATGGCGCCGCGGCAGCATCTGTGCGCGCGAGAAGCTTGCAAACTGGGAAGATCCGC   60
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MetGlyAlaArgAlaSerValLeuSerGIyGIyGluleuAspArgTrpGluIysIleArg  20
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      61 CTGCGCCCCGCGCGCAGAAAGCACTACATGCTGAAGCACTGGTGTGGCCAGCCGCGAG  120
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      21 LeuArgProGIyGIyLysLysLysTyxLysLysLeuLysHisIleValTrrAlaSerArgLu  40
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      121 CTGAGGGGCTTCGCCCTGTAAACCCCGGCTGTGCTGAGAACCCGCGAGGGCTGCAAGCAGATC  180
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      41 LeuGIuaArgPheAlaValaIsnProGIyleuLeuGIuThrSerGIuGIyCyaaRglnIle  60
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      181 ATGAAGCAGCTGCAGCCCGCCCTGCAGACCGGCACCGAGAGACTGCGACCTGTACAAC   240
       : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      61 LeuGIyGlnLeuGIInProSerLeuGIInThrGIySerGIuGIuLeuArgSerLeuTYraSn  80
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      241 ACCGTGGCCACCCTGTACTGCGGTGCACGCCCGGCATCGAGThrCCGCGACACCAAGAGGCC  300
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

D	b	81	ThnValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluVala	100
Q	y	301	CTGACCAAGATTCGAGGAGGAGCAAACTCCAGCAGAAAGCCAGAGGCCCAAGGAG	360
D	b	101	LeuAspLysIleGluGluGlnGlnAsnLysSerLysLysValaGlnGlnAlaVala	120
Q	y	361	GCCGACGGC-----AAGGTAGCCAGAACTACCCCATCTGTGCAGAACTGCAGGGC	411
D	b	121	AspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly	140
Q	y	412	CAGATGTGCACCAAGGCCATCAGCCCCCGCACCTGAACGCTGGGTGAAGTATCGAG	471
D	b	141	GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValaGlu	160
Q	y	472	GAGAAAGCCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCCGACC	531
D	b	161	GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr	180
Q	y	532	CCCCAGGACCTGAACACGATGTGAACACCGTGGCGGCGCACAGGCGCGCATGCAGAT	591
D	b	181	ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMet	200
Q	y	592	CTGAAGGACACCATCAACGAGGAGGCGCGCAGTGGGACCGCTGCACCCCGTCAGAGCC	651
D	b	201	LeuLysGluThrIleAsnGluGlnValaIaIaGluTTrpAspArgValHisProValHisAla	220
Q	y	652	GGCCCCGTGGCCCCCGGCAGATGCGGACCCCCGCGGACGACATGCGCGGCGCAC	711
D	b	221	GlyProIleAlaProGlyGlnMetArgGluProArgLysAspIleAlaGlyThrThr	240
Q	y	712	AGCACCCCTGCAGAGCAGATCGCCTGGATGACCAGCAACCCCCGTCGCGGTGGCGAC	771
D	b	241	SerThrLeuGlnGluGlnIleGlyTTrpMetThrAsnAsnProProIleProValGlyGlu	260
Q	y	772	ATCTACAAGCGGTGATCATCTCTGGGCTTGAACAAGATCTGCGATGTACAACCCCGTG	831
D	b	261	IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValaArgMetLysSerProThr	280
Q	y	832	AGCATCCTTGACATCCGCGCAGAGGCCCAAGAGACCCCTTCGCGACTACGTGAACCGCTTC	891
D	b	281	SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValaAspArgPhe	300
Q	y	892	TTCAAGACCCCTGCGCGCGCAGACAGGCCACCAAGACGTGAAGAACTGATGACCGAGACC	951
D	b	301	TyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTTrpMetThrGluThr	320
Q	y	952	CTGCTGTGTCAGAAAGCCAAACCCGACTGCAAGAACCATCTGCGGCTTCGGCCCCGGC	1011
D	b	321	LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla	340
Q	y	1012	GCCACCCCTGAGAGATGATGATGACCGCTGCCAGGCGGTGGCGGCCCGGCACAAAGGCC	1071
D	b	341	AlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla	360
Q	y	1072	CGCGTGTGCGCGGAGGCGATGAGCCAG--GCCAACAGCGTGAACATCATGATGCAGAA	1128
D	b	361	ArgValLeuAlaGluAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGlnArg	380
Q	y	1129	AGCAACTTCAAGGGCCCCCGCGCAACGTCAAGTGTCTTCAACTGCGGCAAGAGGGCCAC	1188
D	b	381	GlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnCysGlyLysGluGlyHis	400
Q	y	1189	ATCGCCAAAGACTGCGCGCGCCCCCGCAAGAGGGCTGTCTGAAGTGCAGGCAAGAGGGC	1248
D	b	401	ThrAlaArgAsnCysArgAlaProArgLysGlyCysThrTrpLysCysGlyLysGluGly	420
Q	y	1249	CACCAGATGAAGAACTGCACCGAGCGCCAGGCCCAACTTCTTGGGCAAGATCTGGCCAGC	1308
D	b	421	HisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTTrpSer	440
Q	y	1309	CACAAGGGCGCCCCCGGCAACTTCTGTGCAGAAACGCAAGGAGCCCGCGCGCCACCGTG	1368
D	b	441	TyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProPhe	460

QY 1369 -----CCCCACCGCCCCCGCCGAGAGCTTCCGC-----TTCGAG 1404  
DB 461 LeuGlnSerArgProGluProThrAlaProProGluGlnSerPheArgSerGlyValGlu 480  
QY 1405 GAGACCAACCCCGCCCAAGCAGAGCCCAAGACCGCGAGCCCTACCGCGAGCCCTG 1464  
DB 481 ThrThrThrProProGlnLysGlnGluProIleAspLysGluLeuTyr-----ProLeu 498  
QY 1465 ACCGCCCTGCGCAGCCTGTTGCGCAGCGCGCCCTGAGCCAG 1506  
DB 499 ThrSerLeuArgSerLeuPheGlyAsnAspProSerSerGln 512

RESULT 9  
US-08-463-028-8  
; Sequence 8, Application US/08463028  
; Patent No. 6610476  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, Nancy T.  
; APPLICANT: GALLO, Robert C.  
; APPLICANT: WONG-STAL, Flosie  
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Morgan & Flinnegan, L.L.P.  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,028  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 06/693,866  
; FILING DATE: 23-JAN-1985  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 06/659,339  
; FILING DATE: 10-OCT-1984  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Serunian, Leslie A.  
; REGISTRATION NUMBER: 35,353  
; REFERENCE/DOCKET NUMBER: 2026-4193US3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 512 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HTLV-III  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..512  
; OTHER INFORMATION: /note= "gag protein of HTLV-III"  
; US-08-463-028-8

Alignment Scores:  
Pred. No.: 4.99e-128 Length: 512  
Score: 2237.00 Matches: 422  
Percent Similarity: 89.69% Conservative: 39  
Best Local Similarity: 82.10% Mismatches: 39

Query Match: 78.77% Indels: 14  
DB: 4 Gaps: 5  
US-09-475-704A-4 (1-1509) x US-08-463-028-8 (1-512)

QY 1 ATGGCGCCCGCCGCGCAGCATCTGCGCGCGGAGAACTGGAACAATGGAGAAATCCGC 60  
DB 1 MercGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspArgTrpGluLysIleArg 20  
QY 61 CTGGCCCCCGCGCGCAAGAACTACATGCTGAAGCAACCTGTGTGGCCGCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAGGCGCTTCCGCTGAACCCCGCGCTGCTGAGAACCGCGGAGGCTGCAACGATC 180  
DB 41 LeuGluArgPheAlaValAlaAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
QY 181 ATGAAGCAGCTGCAAGCCCGCGCTGAGACCGGCAACGAGAGAGTGCAGCCTGTACAAC 240  
DB 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAAGCCGCGCATGAGGTCCGCAACCAAGAGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100  
QY 301 CTGAACAAGATCGAGAGAGAGCAACAAGTCCAGCAAGAAAGCCAGAGCCCAAGAG 360  
DB 101 LeuAspLysIleGluGluGlnGluAsnLysSerLysLysAlaGlnGlnAlaAla 120  
QY 361 GCCGACGCGC-----AAGTGAGCCAGAACTACCCCATCGTGCAAGCCTGAGGCGC 411  
DB 121 AspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly 140  
QY 412 CAGATGTGCACCAAGCCCATCAGCCCGCGCACCTGAACGCTGGGTGAAGTGAATCGAG 471  
DB 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValValGlu 160  
QY 472 GAGAAAGCCTTCAAGCCCGGAGGTGATCCCATGTTCAACGCGCCTGAGCGGCGCAC 531  
DB 161 GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180  
QY 532 CCCGAGACCTGAACACGATGTTGAACACCGTGGCGCGCCAGCGCCCATGAGATG 591  
DB 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMet 200  
QY 592 CTGAAGACACCATCAACGAGAGAGCGCGCGAGTGGAGACCGCCTGCACCCCGTGAGGCC 651  
DB 201 LeuLysGluThrIleAsnGluGluAlaIleGluTrpAspArgValHisProValHisAla 220  
QY 652 GGGCCCGTGGCGCGCGCGCAGATGCGCGACCCCGCGCGCAGCGACATCGCGCGCCACC 711  
DB 221 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 240  
QY 712 AGCACCTTGACGAGACAGATCGCCTGGATGACCAACCAACCCCGCTGCCGTGGCGGAC 771  
DB 241 SerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsnProProIleProValGlyGlu 260  
QY 772 ATCTACAAGCGGTGATCATCTCTGGGCTGAACAAGATGTCGGATGTACAGCCCGGTG 831  
DB 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280  
QY 832 AGCATCTTGACATCCGCCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAGCCGCTC 891  
DB 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAAGACCTTGGCGCGCGGAGCAGGCGCACCCAGAGCGTGAAGAACTGATGACCGAGACC 951  
DB 301 TyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGluThr 320  
QY 952 CTGCTGCTGACAGAAAGCCCAACCCCGCACTGCAAGACCATCTCGCGGCTCTCGGCCCGGCC 1011  
DB 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340

QY	1012	GCCACCCCTGAGAGGATGATGACCGCCCTGCCAGGCGCTGGGCGGCCCCCGGCCCAAGAGCC	1071
Db	341	AlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlnHisIysAla	360
QY	1072	CCGCGTGTGCGCCGAGCGGATGAGCCAG--GCCAACAGCGCTGAACATCATGATGCAGAAG	1128
Db	361	ArgValLeuAlaGluAlaMetSerGlnValThrAsnThrAlaThrIleMetGlnArg	380
QY	1129	AGCAACTTCAAGGCCCCCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGGGCCAC	1188
Db	381	GlyAsnPheArgAnglnArgIysMetValIysCysPheAsnCySGlyIysGlnGlyHis	400
QY	1189	ATGCGCCAAGAACTGCGCGCGCCCCCGCCAGAGAAGGGCTGCTGAAGTCCGCGCAAGAGGGC	1248
Db	401	ThrAlaArgAsnCysArgAlaProArgIysIysGlyCysTribpCysGlyIysGlnGly	420
QY	1249	CACCAAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATGCGCCAGC	1308
Db	421	HisGlnMetIysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTyrProSer	440
QY	1309	CACAAAGGCGCGCCCGGCAACTTCTGCAGAACCGCAGCGAGCGCGCGCCCGCCACCGTG	1368
Db	441	TyrIysGlyArgProGlnAsnPheLeuGlnSerArgProGluProThrAlaProPhe	460
QY	1369	-----CCCAACGCCCCCGCGCGAGAGCTTCGCG-----TTGAG	1404
Db	461	LeuGlnSerArgProGluProThrAlaProProGluGlnSerPheArgSerGlyValGlu	480
QY	1405	GAGACCAACCCCGCCCCCAAGCAGAGAGCCCAAGAACCGCAGAGCCTTACCGCAGACCCCTG	1464
Db	481	ThrThrThrProProGlnIysGlnGluProIleAspIysGlnIleuTyr-----ProLeu	498
QY	1465	ACCGCCCTGCGCAGCTGTTCGCGAGCGGCGCCCTGAGCCAG	1506
Db	499	ThrSerLeuArgSerLeuPheGlyAsnAspProSerSerGln	512

RESULT 10  
 US-08-816-155B-45  
 Sequence 45, Application US/08816155B  
 Patent No. 5990091  
 GENERAL INFORMATION:  
 APPLICANT: TARTAGLIA, JAMES  
 APPLICANT: COX, WILLIAM I.  
 APPLICANT: GETTIG, RUSSELL R.  
 APPLICANT: MARTINEZ, HECTOR  
 APPLICANT: PAOLETTI, ENZO  
 APPLICANT: PINCUS, STEVEN E.  
 TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
 TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
 NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
 STREET: 745 FIFTH AVENUE  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: USA  
 ZIP: 10151  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/816,155B  
 FILING DATE: 12-MAR-1997  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KOMALSKI, THOMAS J.  
 REGISTRATION NUMBER: 32,147  
 REFERENCE/DOCKET NUMBER: 454310-2990  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-588-0800  
 TELEFAX: 212-588-0500

```

; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
;
US-08-816-155B-45

```

### Alignment Scores:

Pred. No.:	3.51e-127	Length:	50
Score:	2.223.00	Matches:	42
Percent Similarity:	90.16%	Conservative:	38
Best Local Similarity:	82.68%	Mismatches:	36
Query Match:	78.27%	Indels:	14
DB:	2	Gaps:	5

US-09-475-704A-4 (1-1509) x US-08-816-155B-45 (1-500)

QY	1	ATGGGCGCCCGCCAGCATCCTTCGCGCGGCGAGAACTGGACAAGTGGGAGAAGATCCGC	60
Db	1	MctGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTirpGluIleArg	20
QY	61	CTGCGCCCCGGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG	120
Db	21	LeuArgProGlyGlyIleValSerValTyrTyrLeuValHisIleValTirpAlaSerArgGlu	40
QY	121	CTGAGAGGCTTCGCCCCGTGAACCCCGGCTGTGTGAGACCCGCGAGGCTGCAAGCAGATC	180
Db	41	LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle	60
QY	181	ATGAAGCAGCTGCAGCCCGCCCTGCAGACCCGCGAGAGAGCTGGCGCAGCTGTACAC	240
Db	61	LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn	80
QY	241	ACCGTGGCCACCCCTGTACTGCGTGCAGCCGCGCATCGAGGTCCGCGACACCAAGAGGCC	300
Db	81	ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla	100
QY	301	CTGACAAGATCGAGAGAGAGAGACAAGTCCGACGAGAGACCAGAGCCAGGCCAAGGAG	360
Db	101	LeuAspLysIleGluGluGluGlnAsnLysSerTyrLysLysAlaGlnGlnAlaAla	120
QY	361	GCCGACGCGC-----AAGGTGAGCCAGAACTAACCCCATCGTGCAGAACCTGCAGGCGC	411
Db	121	AspThrGlyHisSerAsnGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly	140
QY	412	CAGATGTCACCAAGCCATCAGCCCCCGACACCTGAAAGCCTGGTGAAGTGATCAG	471
Db	141	GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTirpValLysValAlaGlu	160
QY	472	GAGAAAGCCTTCAGCCCCGAGGTGATCCCCCATGTTCAACGCCCTGAGCGAGGCGCCAC	531
Db	161	GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr	180
QY	532	CCCCAGGACCTGAACACGATGTTGAACACCGTGGCGGCGCCACGAGCCGCCATGCAGATG	591
Db	181	ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMet	200
QY	592	CTGAAGGACACCATCAACGAGAGGCGCGCGAGTGGGACCGCTGCACCCCGTGCAGGCC	651
Db	201	LeuLysGluThrIleAsnGluGlnAlaIaGluTirpAspArgValHisProValHisAla	220
QY	652	GGCCCCGTGGCCCCCGGCGAGATGCGGACCCCGCGGCGAGCGACATCGCCGGCGCCAC	711
Db	221	GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr	240
QY	712	AGCACCTGCAGGAGCAGATCGCCTGATGACCAAGCAACCCCGCTGCGCGGCGAC	771
Db	241	SerThrLeuGlnGluGlnIleGlyTirpMetThrAsnAsnProProIleProValGlyGlu	260
QY	772	ATCTACAAGCGGTGATCATCTCGGGCCTGAACAAGATCGTGGGATGTACAGCCCCGTG	831

Db 261 lletyrlvsargtrpilleleleuglyleuasnlysllevalargmettyrserprothr 280  
QY 832 AGCATCTGGACATCCGCCAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTC 891  
Db 281 SerileuaspilleargglnlyprolysgluProphargAsptryValaspargpe 300  
QY 892 TTCAGACCTGCGCGCCGAGCAGGGCCACCGAGACGTGAAGAACTGATGACGAGACC 951  
Db 301 Tyrlsthrleuargalaglnalaserglnlyvallybsanttrymetthrthr 320  
QY 952 CTGCTGTGAGAACGCCCAACCCGACTGCAAGACCATCTGCGCTCGGCCCGCC 1011  
Db 321 LeuleuvalglinsnalaasnProAspCyslysthrilleuylsalaleuglyProala 340  
QY 1012 GCCACCTGGAGAGATGATGACCGCTGCGCAGGGCGGCGGCCGCCCAAGGCC 1071  
Db 341 AlatthrleugluGlumetmetThrAlaCysglnlyValGlylyProglyHlslysla 360  
QY 1072 CGCGTGTGGCGGAGGCGATGAGCCAG---GCCAACAGCGTGAACATCATGATGACAGAG 1128  
Db 361 ArgvalleuvalaglualametserglnvalThrAsnseralathrilemetmetglnarg 380  
QY 1129 AGCAACTTCAAGGGCCCCCGCGCAACGTCAAGTGTCAACTGCGGCAAGAGGGCCAC 1188  
Db 381 GlyAsnphearGAsnglnarglysllevallyscyspheasnCysglylysgluGlyHls 400  
QY 1189 ATGCCAAGAACTGCGCGCCCCCGCGCAAGAGGGCTGTGAACTGCGGCAAGAGGGGC 1248  
Db 401 ThrAlaargAsnCysargAlaProarglysllyCyslysttrylsCysglylysgluGly 420  
QY 1249 CACCAGATGAAGACTGACACCGAGCGCGCAAGCCCACTTCTGGGCAAGATCTGGCCAGC 1308  
Db 421 HisglnmetlysaSpCysThrGluargGlnAlaAsnphleuGlylyslletrProser 440  
QY 1309 CACAAGGGCGCGCCCGCAACTTCTGCGAAGACCGCAGCGAGCCCGCGCCCAACCGTG 1368  
Db 441 TyrlsglylArgProglyAsnphleuGlnserArgProglu----- 454  
QY 1369 CCCACCGCCCCCGCGCGAGAGCTTCCGC-----TTGAGAGACCAACCCCGCCCC 1422  
Db 455 ProthrAlaProProgluGluSerPheargserglyValglutThrThrProProglu 474  
QY 1423 AAGCAGAGCCCAAGAGCCGAGCCCTTACCGGAGCCCTGACCGCCTCGCAGCCTG 1482  
Db 475 LysglnGluProileaspIysgluLeuTyrr-----ProleuthrserleuArgserleu 492  
QY 1483 TTCGGCAGCGGCGCCCTGAGCCAG 1506  
Db 493 PheglyAsnaspProserSerGln 500

RESULT 11  
US-09-079-587-45  
; Sequence 45, Application US/09079587  
; Patent No. 6130066  
; GENERAL INFORMATION:  
; APPLICANT: TARTAGLIA, JAMES  
; APPLICANT: COX, WILLIAM I.  
; APPLICANT: GETTIG, RUSSELL R.  
; APPLICANT: MARTINEZ, HECTOR  
; APPLICANT: PAOLETTI, ENZO  
; APPLICANT: PINCUS, STEVEN E.  
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
; TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
; STREET: 745 FIFTH AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10151  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,587  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/816,155  
FILING DATE: 12-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: KOWALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2990  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 amino acids  
TYPE: amino acid  
STRANDEDNESS: n/a  
TOPOLOGY: linear  
MOLECULE TYPE: amino acid  
US-09-079-587-45

Alignment Scores:  
Pred. No.: 3,51e-127 Length: 500  
Score: 2223.00 Matches: 420  
Percent Similarity: 90.16% Conservative: 38  
Best Local Similarity: 82.68% Mismatches: 36  
Query Match: 78.27% Indels: 14  
DB: 3 Gaps: 5

US-09-475-704A-4 (1-1509) x US-09-079-587-45 (1-500)

QY 1 ATGGCGCGCGCGCCAGCATCTGCGCGGAGAACTGACAACTGAGAGATCCGC 60  
Db 1 MetGlyAlaargAlaSerValleuSerGlyGlyGluLeuAspArgTrpGluysilearg 20  
QY 61 CTGGCGCGCGCGGCAAGAACTACATGCTGAAGCACTGTGTGGCGCGCGAG 120  
Db 21 LeuArgProgllyGlylysllysttryllyleuLysHlsileValTrpAlaserArglu 40  
QY 121 CTGAGAGGCTTGCCTGAACCCCGCGCTGTGAGACCGCGGAGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaValaAsnProglyLeuLeuGluThrserGluGlyCysArgGlnle 60  
QY 181 ATGAAGCAGCTGACCGCGCGCTGACAGCCGCAACCGAGAGCTGCGAGCTGTACAAC 240  
Db 61 LeuGlyGlnleuGlnProserleuGlnThrGlyserGluGluLeuArgserleuTyrrasn 80  
QY 241 ACCGTGGCCACCTGTACTGCGGTGACGCGCGCATGAGGTCCGCGACCAAGAGGCC 300  
Db 81 ThrValAlaThrleuTyrrCysValHlsGlnArglleGluileLysAspThrlysgluAla 100  
QY 301 CTGACAAAGATCGAGAGAGAGCAACAAGTCCAGCAGAGAACCCAGAGCCCAAGAG 360  
Db 101 LeuAspIysileGluGluGlnAsnlyserLyslysllyslalaglnGlnAlaAla 120  
QY 361 GCCGACGCG-----AAGGTGAGCCAGAACTACCCCATCGTGAGAACTGACAGGC 411  
Db 121 AspThrGlyHlsSerAsnGlnValserGlnAsnTyrrProleValGlnAsnilleGlnly 140  
QY 412 CAGATGTGACCAAGCCATACGCCCGCGACCTGAAAGCGCTGAGGTGATGATGAG 471  
Db 141 GlnMetValHlsGlnAlaIleserProArgThrleuAsnAlaTrpValysValValGlu 160  
QY 472 GAGAAGGCTTACGCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGGCGGCCACC 531  
Db 161 GlulysAlaPheSerProgluValillePrometPheSerAlaLeuSerGluGlyAlaThr 180  
QY 532 CCCGAGCACTGAACAGATGTTGAACACCGTGGCGGCGCACAGCGCCCATGAGATG 591

Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyHisGlnAlaIleMetGlnMet 200  
QY 592 CTGAAGGACACCATCAACGAGAGGCGCGAGTGGACCGCTGCACCCCGTGCAGGCC 651  
Db 201 LeuYsgluThrIleAsnGluGlnAlaIleGluThrPaspArgValHisProValHisAla 220  
QY 652 GGCCCGGTGGCCCCCGGCGAGATGCGCGACCCCGCGGCGAGCATCGCGCGGCCACC 711  
Db 221 GlyProIleAlaProGlyGlnMetArgIleuProArgIleSerAspIleIleGlyThrThr 240  
QY 712 AGCACCTTGACGAGACGATCGCTGATGACCAACCCCGCTGCGCGTGGCGAC 771  
Db 241 SerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProIleProValGlyGlu 260  
QY 772 ATCTACAGCGGTGATCATCTGGCTGAAACAAGATGTCGGATGTACAGCCCGTG 831  
Db 261 IleTyrlsArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrlSerProThr 280  
QY 832 AGCATCTTGACATCCGCGACGAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTC 891  
Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrlValAspArgPhe 300  
QY 892 TTCAAGACCTTGCGCGCGGCGAGCAGGCCACCCAGACGCTGAGAACTGATGACCGAGACC 951  
Db 301 TyrlsThrLeuArgAlaGluGlnAlaIleSerGlnGluValLysAsnTrpMetThrGluThr 320  
QY 952 CTGCTGTGTCAGAACGCCCAACCCGACTGCAAGACCATCTCGCGCTTCGCGCGCGC 1011  
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysIleThrIleLeuLysAlaLeuGlyProAla 340  
QY 1012 GCCACCTTGAGAGAGATGATGACCGCTGCCAGGGCGGTGGCGGCCCAAGGCC 1071  
Db 341 AlaThrLeuGlnGluGlnMetMetThrAlaCysGlnGlyValGlyProGlyHisLysAla 360  
QY 1072 CGCGTGTGGCGGAGGCGCATGAGCCAG---GCCAACAGCGTGAACATCATGTCAGAAAG 1128  
Db 361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArg 380  
QY 1129 AGCAACTTCAAGGGCGCGCGCAACGTCAGTGTCTCAACTGCGGCAAGAGGGCCAC 1188  
Db 381 GlnAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCysGlyLysGluHis 400  
QY 1189 ATCGCCCAAGAACTGCGCGCGCGCGCGCAAGAGGGCTGTGAACTGCGGCAAGAGGGC 1248  
Db 401 ThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGly 420  
QY 1249 CACCAAGATGAAGACTGCAACGAGCGCGCAAGGCCAACTTCTGGGCAAGATCTGGCCACG 1308  
Db 421 HisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440  
QY 1309 CACAAAGGGCGCGCGCAACTTCTGTGAGAAACCGCAGCGAGCCCGCGCCCAACCTG 1368  
Db 441 TyrlsGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 454  
QY 1369 CCCACGCGCGCGCGCGCGAGAGCTTCGCGC-----TTCGAGGAGACCAACCCCGCGCC 1422  
Db 455 ProThrAlaProProGluGlnSerPheArgSerGlyValGluThrThrProProGln 474  
QY 1423 AAGCAGAGAGCCCAAGAGCGCGAGCCCTTACCGCGAGCCCTGACCGCGCTGCGAGCCTG 1482  
Db 475 LysGlnGluProIleAspLysGluLeuTyrl-----ProLeuThrSerLeuArgSerLeu 492  
QY 1483 TTCGGCAGCGCGCGCGCTGAGCCAG 1506  
Db 493 PheGlyAsnAspProSerSerGln 500

RESULT 12  
US-09-952-060-33  
; Sequence 33, Application US/09952060  
; Patent No. 6733993  
; GENERAL INFORMATION:  
; APPLICANT: Emimil, Emilio A.

APPLICANT: Youil, Rima  
APPLICANT: Betc, Andrew J.  
APPLICANT: Chen, Ling  
APPLICANT: Kaslow, David C.  
APPLICANT: Shiver, John W.  
APPLICANT: Toner, Timothy J.  
APPLICANT: Casimiro, Danilo R.  
TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  
TITLE OF INVENTION: MODIFICATIONS  
FILE REFERENCE: 20747Y  
CURRENT APPLICATION NUMBER: US/09/952,060  
CURRENT FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: PCT/US01/28861  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/317,814  
PRIOR FILING DATE: 2001-09-07  
PRIOR APPLICATION NUMBER: 60/279,056  
PRIOR FILING DATE: 2001-03-27  
PRIOR APPLICATION NUMBER: 60/233,180  
PRIOR FILING DATE: 2000-09-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 33  
LENGTH: 493  
TYPE: PRT  
ORGANISM: Artificial Sequence  
OTHER INFORMATION: tPA-gag fusion open reading frame  
US-09-952-060-33  
Alignment Scores:  
Pred. No.: 2.29e-119 length: 493  
Score: 2094.00 Matches: 396  
Percent Similarity: 90.53% Conservative: 34  
Best Local Similarity: 83.37% Mismatches: 31  
Query Match: 73.73% Indels: 14  
DB: 4 Gaps: 5  
US-09-475-704A-4 (1-1509) x US-09-952-060-33 (1-493)  
QY 100 CTGCTGTGGGCGCAGCGCGAGCTGAGGGCTTCGCGCTGAACCCCGGCTGTGAGACC 159  
Db 27 ILevalITPAlaSerArgGluLeuGluArgPheAlaValAsnProGlyLeuLeuGluThr 46  
QY 160 GCCGAGGGCTGCAAGCAGATCATGAAGCAGCTGCAGCGCGCGCTGCAGACCGGACGAG 219  
Db 47 SerGluGlyCysArgGlnIleLeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlu 66  
QY 220 GAGCTGCGCAGCCTGTACAACACCGTGGCCACCTGTACTGCGTGACGCGCGCATGAG 279  
Db 67 GluLeuArgSerLeuTyrlAsnThrValAlaThrLeuTyrlCysValHisGlnLysIleAsp 86  
QY 280 GTCCGCGACACCAAGAGAGCGCGCTGCAAGATCGAGAGGAGCAGACAAGTCCAGCAG 339  
Db 87 ValLysAspThrLysGlnAlaLeuGluLysIleGluGluGlnAsnLysSerLysLys 106  
QY 340 AAGACCCAGCAGCGCAAGAGAGCGCGCGC-----AAGGTGACCGAAGACTTACCC 390  
Db 107 LysAlaGlnGlnAlaAlaIleGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrlPro 126  
QY 391 ATCGTCAGAACTGAGAGGCGCAGATGTGTGACACGAGCCATCAGCCCGCACCTGAGAC 450  
Db 127 IleValGlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsn 146  
QY 451 GCCTGGGTGAAGGTGATCGAGAGAGGCGCTTACGCCCGGAGGTGATCCCATGTTCACC 510  
Db 147 AlaTrpValLysValValGluGluLysAlaPheSerProGluValIlePrometPheSer 166  
QY 511 GCCCTGAGCAGGCGCGCACCCCGCAGACCTGAACAAGATGTGAACAACCGTGGCGGC 570  
Db 167 AlaLeuSerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGly 186

QY 571 CACCAGGCGCCATGACATGCTGAGGACACATCAACGAGAGGCGCCGAGTGAGAC 630  
Db 187 HlGlnAlaAlaMetGlnMetLeuLysGlnThrIleAsnGlnGlnAlaAlaGlnTrpAsp 206  
QY 631 CGCCTGACCCCGTGACAGGCGGCGCCGTCGCGCCCGGCGAGATGCGGACCCCGCGCG 690  
Db 207 ArgLeuHisProValHisAlaGlyProIleAlaProGlyGlnMetArgGlnProArgGly 226  
QY 691 AGCGACATCGCCGCGCGCCACGACACCTGCGAGAGACAGATCGCCTGATGACGACCAAC 750  
Db 227 SerAspIleAlaGlyThrThrSerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsn 246  
QY 751 CCCCCCGTGGCGGTGGGCGACATCTACAAGCGGTGATCATCTTGCGCTGAACAAGATC 810  
Db 247 ProProIleProValGlyGlnIleTyxLysArgTrpIleIleLeuGlyLeuAsnLysIle 266  
QY 811 GTGCGGATGTACAGCCCGCGTGCATCTGACATCCGCCAGGCGCCCAAGAGCCCTTC 870  
Db 267 ValArgMetTySerProThrSerIleLeuAspIleArgGlnGlyProLysGlnProPhe 286  
QY 871 CGCGACTACGTGACCGCTTCTTCAAGACCTTGCAGCGCGCGGCGAGAGGCGCAACGAGACGTG 930  
Db 287 ArgAspTyxValAspArgPheTyxLysThrLeuArgAlaGlnAlaSerGlnGlnVal 306  
QY 931 AAGAACTGGATGACCGGAGACCTGCTGTGTCAGAACGCCAACCCCGACTGCAAGACCTTC 990  
Db 307 LysAsnTrpMetThrGlnThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIle 326  
QY 991 CTGCGCGCTCTCGGCGCGCGCGCGCACCTTGAGAGATGATGACCGCTGCGAGGCGGTG 1050  
Db 327 LeuLysAlaLeuGlyProAlaAlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyVal 346  
QY 1051 GCGGCGCGCGCGCGCACAGGCGCGGTGTGCGCGGAGGCGATGAGCCAG--GCCAACAGC 1107  
Db 347 GlyGlyProGlyHisLysAlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnSer 366  
QY 1108 GTGAACATCATGATGACAGAGAGCAACTTCAAGGCGCGCGCGCAAGCTCAAGTGCTTC 1167  
Db 367 AlaThrIleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPhe 386  
QY 1168 AACTGCGGCAAGAGGCGCACATCGCCAGAACTGCGCGCGCGCGCGCAAGAGGCGTGC 1227  
Db 387 AsnCyseGlyLysValGlyHisIleAlaLysAsnCyseArgAlaProArgLysGlyCys 406  
QY 1228 TGGAGTGGCGGCAAGAGGCGCGCACGATGAAGACTGCAACGCGCGCGCAACTTC 1287  
Db 407 TrpLysCyseGlyLysGlnGlyHisGlnMetLysAspCysAsnGlnArgGlnAlaAsnPhe 426  
QY 1288 CTGGGCAAGATCTGGCGCGCGCACAGGCGCGCGCGCAACTTCTGCAAGACCGCAGC 1347  
Db 427 LeuGlyLysIleTrpProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgPro 446  
QY 1348 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGCTTCCGCTTC----- 1401  
Db 447 Gln-----ProThrAlaProProGlnGlnSerPheArgPheGlyGln 460  
QY 1402 GAGGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1461  
Db 461 GlnLysThrThrProSerGlnLysGlnGlnProIleAspLysGlnLeuTyx-----Pro 478  
QY 1462 CTGACCGCGCGCTGCGCGAGCTGTTCGCGAGCGGCGCGCGCGCGCGCGCGCGCGCGC 1506  
Db 479 LeuAlaSerLeuArgSerLeuPheGlyAsnAspProSerSerGln 493

RESULT 13  
US-08-127-499A-11

; Sequence 11, Application US/08127499A  
; Patent No. 5510264  
; GENERAL INFORMATION:  
; APPLICANT: VAN ALSTYNE, Diane  
; APPLICANT: SHARMA, Lawrence Rajendra  
; TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED  
; TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES

NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/127,499A  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 51916/102/INBI  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 478 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
US-08-127-499A-11

Alignment Scores:  
Pred. No.: 8,57e-115 Length: 478  
Score: 2018.50 Matches: 382  
Percent Similarity: 87.86% Conservative: 45  
Best Local Similarity: 78.60% Mismatches: 40  
Query Match: 71.07% Indels: 19  
DB: 1 Gaps: 5

US-09-475-704A-4 (1-1509) x US-08-127-499A-11 (1-478)

QY 1 ATGGGCGCGCGCGCGCACATCTGCGCGGAGAGCTGGAAGTGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspArgTrpGlyLysIleArg 20  
QY 61 CTGGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGCGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysTyxLeuLeuLysHisIleValTrpAlaSerArgGln 40  
QY 121 CTGAGGCGCTTCCGCTGAACCCCGCGCTGAGAGACCGCGAGGCGTGCAGACAGATC 180  
Db 41 LeuGlnArgPheAlaValAsnPProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60  
QY 181 ATGAAGCAGCTGCAGCGCGCGCTGACAGACCGGACCGAGAGCTGCGAGCTGTACAAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnLeuArgSerLeuTyxAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGCGCGCGCATGAGGTCCGCGACCAAGAGGCGC 300  
Db 81 ThrValAlaThrLeuTyxTyxValHisGlnArgIleGlnIleLysAspThrLysGlnAla 100  
QY 301 CTGACAAAGATCGAGAGAGAGAGCAAGCAAGTCCAGCAGAGAACCCAGAGCCCAAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnGlnAsnLysSerLysLysLysAlaGlnGlnAlaAla 120  
QY 361 GCCGACGCGC-----AAGTGAGCGCAAGCACTACCCCATGCTGCAGAACCTGACGCGC 411  
Db 121 AspThrGlyHisSerSerGlnValSerGlnAsnTyxProIleValGlnAsnIleGlnGly 140  
QY 412 CAGATGTGCACGAGCGCATCAAGCCCGCGACCCCTGAAGCCTGGTGAAGGTGATCGAG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValGln 160



Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysAlaGlnAlaIaIa 120  
QY 361 GCCGACGCGC-----AAGGTGAGCCAGAACTACCCCATCTGTGAGAACCTGACGGGC 411  
Db 121 AspThrGlyHisSerSerGlnValSerGlnAsnTyProIleValGlnAsnIleGlnGly 140  
QY 412 CAGATGTTGACACGAGCCATCAGCCCCCGACACCTGAAACGCTTGAGTGAAGTGTGAG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValGlu 160  
QY 472 GAGAAGGCTTACGCGCGAGGTGATCCCATGTTACCGCCCTGAGCGAGGCGCCACC 531  
Db 161 GluLysAlaPheSerProGluValIlePrometPheSerAlaLeuSerGluGlyAlaThr 180  
QY 532 CCCGAGGACCTGAACACGATGTTGAACACCGTGGCGGCCACGAGCCCGCATGAGATG 591  
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIaMetGlnMet 200  
QY 592 CTGAAGGACACCATCAACGAGAGGCGCGGAGTGAGACCGCTGCACCCCGTGAAGGCC 651  
Db 201 LeuLysGluThrIleAsnGluGlnAlaIaGluTyrAspArgValHisProValHisAla 220  
QY 652 GGGCCCGTGGCGCGCGGCGAGATGGCGGACCGCGCGGCGGACATCGCGCGGCCACC 711  
Db 221 GlyProIleAlaProGlyGlnMetCargLupProArgGlySerAspIleAlaGlyThrThr 240  
QY 712 AGCACCTTGACAGACGATCGCTGATGACCAACCCCGCTGCGCGCGGCGGAC 771  
Db 241 SerThrLeuGlnGluGlnIleGlyTyrMetThrAsnAsnProIleProValGlyGlu 260  
QY 772 ATCTACAAGCGGTGATCATCTCTGGGCTGAAACAGATCGTGGGATGTACAGCCCGTG 831  
Db 261 IleTyrLysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280  
QY 832 AGCATCTTGACATCCGCCAGGCGGCCAACGAGACCTTCCGCGACTACGTGACCGCTTC 891  
Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAAGACCTGCGCGCGCGAGGAGGCCAACGAGACGTGAAGAACTGATGACCGAGACC 951  
Db 301 TyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTyrMetThrGluThr 320  
QY 952 CTGCTGTGACAGACGCCAACCCCGACTGCAAGACCATCTCGCGGCTCTCGGCGCGCGC 1011  
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340  
QY 1012 GCCACCTTGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCGCGGCCAACAGGCC 1071  
Db 341 AlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla 360  
QY 1072 CGCGTGTGGCGGAGGCGATGAGCCAG--GCCAACAGCGTGAACATCATGTGACAGAG 1128  
Db 361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGlnArg 380  
QY 1129 AGCAACTTCAAGGCGCGCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGGCCAC 1188  
Db 381 GlnAsnPheArgAsnGlnArgLysMetValLysCysPheAsnCysGlyLysGluGlyHis 400  
QY 1189 ATGCGCAAGACTGCGCGCGCGCGCGCAAGAGGGCTGCTGAGAAGTGGCGCAAGAGGGC 1248  
Db 401 ThrAlaArgAsnCysArgAlaProArgLysGlyCysTyrLysCysGlyLysGluGly 420  
QY 1249 CACGAGATGAAGACTGACCGGAGCGCGGCAACTTCTCGGCGCAAGATC---TGCGCC 1305  
Db 421 HisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleCysLysLeuPro 440  
QY 1306 AGCCACAAGGCGCGCGCGCAACTTCTGCAAGACCGGAGGAGCGCGCGCGGCCACC 1365  
Db 441 ThrArgGluGlyGlnGlyIlePhePheArgAlaAspGlnSerGln----- 455  
QY 1366 GTGCCCAACGCGCGCGCGAGAGCTTTCGGCTTGAAGAGACCAACCGCGCGCGGCCAAG 1425

Db 456 -----GlnProHisHisPhePheArgAlaAspGlnSer----- 466  
QY 1426 CAGAGCCCAAGACCGC 1443  
Db 467 GlnGlnProHisGlnLys 472  
RESULT 15  
US-08-392-794A-2  
; Sequence 2, Application US/08392794A  
; Patent No. 6025141  
; GENERAL INFORMATION:  
; APPLICANT: HU, Yu-wen  
; TITLE OF INVENTION: IMMUNOFLUORESCENCE ASSAY FOR THE DETECTION OF  
; TITLE OF INVENTION: ANTIBODIES USING RECOMBINANT ANTIGENS IN INSOLUBLE FORM  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/392,794A  
; FILING DATE: 09-DEC-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/164,789  
; FILING DATE: 10-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEIMARK, SHERIDAN  
; REGISTRATION NUMBER: 20,520  
; REFERENCE/DOCKET NUMBER: HU=4A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 437 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-392-794A-2  
Alignment Scores:  
Pred. No.: 2.41e-114 Length: 437  
Score: 2011.00 Matches: 374  
Percent Similarity: 93.14% Conservative: 33  
Best Local Similarity: 85.58% Mismatches: 26  
Query Match: 70.81% Indels: 4  
DB: 3 Gaps: 2  
US-09-475-704A-4 (1-1509) x US-08-392-794A-2 (1-437)  
QY 1 ATGGGCGCGCGCGGACGATCTGCGCGGAGAGACTGAGAGTGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTyrGluLysIleArg 20  
QY 61 CTGCGCGCGCGGCGGCAAGACACTACGTGTAAGCACCTGTGTGGCGGAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysTyrLysLysLysHisIleValTyrAlaSerArgGlu 40  
QY 121 CTGAGGCGCTTGGCGCTGAACCCCGGCGCTGTGAGAGACCGCGGAGCTGCAAGAGATC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
QY 181 ATGAAGCAGCTGACGCGCGCGCTGAGACCGGACCGAGAGAGTGGCGAGCTGTACAAC 240

```
Db      61 LeuGlnLeuGlnProSerLeuGlnThrGlySerGlnGluLeuMetSerLeuTyrAsn 80
QY      241 ACCGTGGCCACCCCTGTACTGCGTGACGCCGACATCGAGGTCCGGACACCAAGAGGCC 300
      81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleTyrAspThrLysGluAla 100
QY      301 CTGACAAGATCGAGAGAGACAGAACAGTCCACAGACAGAACCCAGACAGCCAAAGAG 360
      101 LeuAspLysIleGlnGlnGlnGlnLysSerLysLysAlaGlnGlnAlaAla 120
QY      361 GCCGACGGC-----AAGGTAGCCAGAACTACCCCATCTGTGAGAACCTGACGGGC 411
      121 AspThrGlnHisSerAsnGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly 140
QY      412 CAGATGGTGACACCAAGCCATCAGCCCCCGACACCTGAAAGCCTGGGTGAAGTGAATCGAG 471
      141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTPrValLysValValGlu 160
QY      472 GAGAAAGCCTTCAGCCCCCGAGGTGATCCCATGTTCACCGCCCTGACCGAGGGCCACC 531
      161 GluLysAlaPheSerProGluValIlePrometPheSerAlaLeuSerGlnGlyAlaThr 180
QY      532 CCCCAGGACCTGAACACGATGTTGAACACCGTGGGGGGCCACCAAGCCGCCATGCAATG 591
      181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMet 200
QY      592 CTGAAGGACACCATCAACGAGAGAGCGCCGAGTGGGACCGCCCTGACACCCCTGACAGGC 651
      201 LeuLysGlnThrIleAsnGlnGlnAlaIleGluTPrAspArgValHisProValHisAla 220
QY      652 GGGCCCCGTGGCCCCCGCCAGATGGCGGACACCCCGCGGACAGCGACATCGCCGGCCACC 711
      221 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 240
QY      712 AGCACCTTGACGAGACAGATCGCCTGAGATGACCAAGACACCCCGCTGCCGTGGCGGAC 771
      241 SerThrLeuGlnGlnGlnIleGlyTPrMetThrAsnAsnProProIleProValGlyGlu 260
QY      772 ATCTACAAGCGGTGATCATCTCGGCTGAAACAAGATCGTCCGATGTACAGCCCGCTG 831
      261 IleTyrLysArgTPrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280
QY      832 AGCATCTTGACATCCGCCAGGAGGCCCAAGAGGCCCTTCCGGACTACGTGACCGCTTC 891
      281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300
QY      892 TTCAAGACCCCTGCGCGCCGAGACAGGCCACCCAGAGCGTGAAGACTGGATGACCGAGACC 951
      301 TyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTPrMetThrGluThr 320
QY      952 CTGCTGGTGCAAGCGCCAAACCCGACTGCAAGACCATCTGCGGCTCTGGCCCGGC 1011
      321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340
QY      1012 GCCACCCCTGAGGAGATGATGACCCCTGACAGGGCGTGGCGGCCCGGCCACAAGGCC 1071
      341 AlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlnHisLysAla 360
QY      1072 CGCGTGTGGCCGAGCGATGAGCCAG--GCCAACAGCGTGAACATCATGATGCAAGAAG 1128
      361 ArgValIleuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArg 380
QY      1129 AGCAACTTCAAGGCGCCCGCGCAACGTCAGTGCTCAACTGCGGCAAGAGAGGCCAC 1188
      381 GlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCysGlyLysGlnGlnHis 400
QY      1189 ATGCCCAAGAACTGCGCGCGCCCGCCCAAGAGGCTGCTGAAGTGCAGGAGAGGCC 1248
      401 ThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTPrLysCysGlyLysGlnGly 420
QY      1249 CACCAAGATGAAGACTGACCGGACCGCCAGGCCAACTTCTGGGCAAGATC 1299
```

```
Db      421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIle 437
..
Search completed: March 11, 2005, 15:46:26
Job time : 76.6742 secs
```

**THIS PAGE BLANK (USE 19)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 14:56:14 ; Search time 206.806 Seconds  
(without alignments)  
5644.151 Million cell updates/sec

Title: US-09-475-704A-4  
Perfect score: 2840  
Sequence: 1 atgagcgccgcgcgcagcat.....gcgagccctgagccagctaa 1509

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues  
Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool\_p/US09475704/runat.10032005.140220.14808/app.query.fasta.1.3342  
-DB=A\_Geneseq\_16Dec04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704.@CGN\_1.1\_472.@runat.10032005.140220.14808 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2661	93.7	502	3	AAy96944	AAy96944 HIV synch
2	2464	86.8	492	6	AAE37600	AAE37600 HIV-1 sub
3	2463	86.7	492	8	ADP02848	ADP02848 wild type
4	2463	86.7	513	8	ADP02847	ADP02847 wild type
5	2457	86.5	492	5	AAE18303	AAE18303 Human imm
6	2457	86.5	635	5	AAm48948	AAm48948 HIV-1 sub
7	2440	85.9	492	7	ADC72879	ADC72879 HIV-1 gag
8	2430	85.6	492	7	ADC72880	ADC72880 HIV-1 gag
9	2421.5	85.3	491	7	ADC72910	ADC72910 HIV-1 gag
10	2415.5	85.1	1457	8	ADN36406	ADN36406 Human pro

11	2413.5	85.0	491	7	ADC72913	ADC72913 HIV-1 gag
12	2410.5	84.9	514	8	ADN36396	ADN36396 Human pro
13	2404	84.6	498	3	AAB69276	AAB69276 HIV-1 non
14	2402	84.6	494	3	AAB69275	AAB69275 HIV-1 non
15	2387	84.0	508	7	ADC72875	ADC72875 HIV-1 gag
16	2385.5	84.0	491	7	ADC72874	ADC72874 HIV-1 gag
17	2381.5	83.9	491	3	AAB69278	AAB69278 HIV-1 non
18	2375.5	83.6	497	7	ADC72877	ADC72877 HIV-1 gag
19	2353.5	82.9	487	7	ADC72876	ADC72876 HIV-1 gag
20	2348	82.7	496	7	ADC72878	ADC72878 HIV-1 gag
21	2343	82.5	492	3	AAy96943	AAy96943 HIV synch
22	2328.5	82.0	3025	4	AAB69271	AAB69271 HIV-1 sub
23	2273	80.0	494	3	AAB69271	AAB69271 HIV-1 non
24	2270	79.9	500	1	AAp80884	AAp80884 Sequence
25	2258.5	79.5	1229	6	AAE37597	AAE37597 HIV-1 Grt
26	2258	79.5	500	7	ADC72788	ADC72788 HIV-1 gag
27	2253	79.3	937	2	AAE36822	AAE36822 PE bindin
28	2252.5	79.3	497	2	AAE36822	AAE36822 PE bindin
29	2251	79.3	442	6	AAE37598	AAE37598 Sequence
30	2247.5	79.1	494	7	ADC72896	ADC72896 HIV-1 gag
31	2247	79.1	502	2	AAW53108	AAW53108 Gag prote
32	2247	79.1	502	3	AAy77294	AAy77294 HIV-1 (AT
33	2247	79.1	502	6	ABU63178	ABU63178 Protein #
34	2247	79.1	506	1	AAp60176	AAp60176 Sequence
35	2247	79.1	506	2	AAE37578	AAE37578 ARV-2 (9B
36	2247	79.1	506	5	AAE37578	AAE37578 HIV-1 gag
37	2246	79.1	500	2	AAW63737	AAW63737 HIV-1 gag
38	2246	79.1	500	3	AAAB10046	AAAB10046 HIV-1 gag
39	2246	79.1	500	3	AAy70599	AAy70599 Codon opt
40	2246	79.1	500	3	AAy70598	AAy70598 wild type
41	2246	79.1	500	6	ABP96542	ABP96542 HIV-1 NYU
42	2246	79.1	500	6	ABR82426	ABR82426 HIV Gag p
43	2246	79.1	500	7	ADJ92015	ADJ92015 Human imm
44	2246	79.1	500	8	AD128568	AD128568 HIV-1 gag
45	2246	79.1	512	8	AD128569	AD128569 HIV-1 gag

ALIGNMENTS

RESULT 1	AAy96944	standard; protein; 502 AA.
ID	AAy96944;	
AC	AAy96944;	
DT	31-OCT-2000	(first entry)
XX		
DE	HIV synthetic Gag containing polypeptide.	
XX		
KW	Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;	
KW	DNA immunization; packaging cell line; antigen presentation.	
XX		
OS	Human immunodeficiency virus; type C strain AF110967.	
OS	Synthetic.	
XX		
PN	W0200039304-A2.	
XX		
PD	06-JUL-2000.	
XX		
PF	30-DEC-1999;	99WO-US031273.
XX		
PR	31-DEC-1998;	98US-0114495P.
PR	01-SEP-1999;	99US-0152195P.
XX		
PA	(CHIR ) CHIRON CORP.	
XX		
PI	Barnett S, Zur Megede J;	
XX		
DR	WPI; 2000-452401/39.	
XX		
PT	Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env	
PT	polypeptide and the polypeptide useful for immunizing a mammal especially	
PT	human against HIV.	

XX Example 1; Page 105-106; 113pp; English.  
XX  
XX Expression cassettes comprising a polynucleotide encoding antigenic type  
CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
CC in DNA immunization, generation of packaging cell lines and production of  
CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
CC cassettes exhibit increased potency for induction of cytotoxic T-  
CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-  
CC assemble into non-infectious virus-like particles which are used as a  
CC matrix for the proper presentation of an antigen entrapped or associated  
CC to the immune system of the host  
XX  
SQ Sequence 502 AA;

Alignment Scores:  
Pred. No.: 1.76e-143 Length: 502  
Score: 2661.00 Matches: 502  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 93.70% Indels: 0  
DB: 3 Gaps: 0

US-09-475-704A-4 (1-1509) x AAY96944 (1-502)

QY 1 ATGGGCGCCCGCGCCGACATCTGCGCGCGGAGAGCTGGACAAGTGGAGAGAATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyIleuAspLysTrpGlyLysIleArg 20  
QY 61 CTGCGCGCCCGCGGAGAGAGCACTATGCTGAAGCACTGCTGTGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGGGGCTTCCGCTGAACCCCGGCTGCTGAAGACCGCGGAGGCTGCAAGCATC 180  
Db 41 LeuGlnGlyPheAlaLeuAsnProGlyLeuLeuGluThrAlaGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGCAAGCCCGCCCTGAGACCGGCAAGGAGAGCTGGCAGCTGTACAAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGCGCGGCACTGAGGTCCGCGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAAGAGAGAGACAAGTCCGACGACAGAGACCAAGAGAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120  
QY 361 GCCGACGCGCAAGGTGAGCCAGAACTACCCCATCTGTCAGAGACCTGCAAGGCGCAGTGTG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMetVal 140  
QY 421 CACCAAGCCATCAGCCCGCCGACCCCTGAACGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLysAla 160  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCGCAGGAC 540  
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 180  
QY 541 CTGAACAAGATGTTGAACACCGGTGGGGCGCCACAGCGCGCCATGCAAGTGTGAAGGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGGAGCGCGCTGCAACCCGTCAGAGCGCGCGCTG 660  
Db 201 ThrIleAsnGlnGluAlaAlaGlnTrpAspArgLeuHisProValGlnAlaGlyProVal 220  
QY 661 GCCCGCGCGCAAGTGGCGGAGCCCGCGCGGCAAGCATCGCGCGCGCCACCAAGCACCTG 720  
Db 221 AlaProGlyGlnMetArgAspProArgGlySerAspIleAlaGlyAlaThrSerThrLeu 240

QY 721 CAGAGCAGATCGCTGATGACCAACCCCGCGTGCCTGGGCGGACATCTACAAG 780  
Db 241 GlnGlnGlnIleAlaTrpMetThrSerAsnProProValProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCTGTGAACAAGATCGTGGGATGTAGACCCCGTGAATCCTG 840  
Db 261 ArgTrpIleIleuGlnLysLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGAGGCGCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTCAAGACC 900  
Db 281 AspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPheLeuThr 300  
QY 901 CTGCGCGCCGAGAGAGCCACCAAGAGCGTGAAGAACTGATGACCGAGACCTGTGTG 960  
Db 301 LeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrGluThrLeuVal 320  
QY 961 CAGAAGCCCAACCCGACTGCAAGACCATCTGCGCGCTGCGCCCGCGGCAACCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340  
QY 1021 GAGAGATGATGACCGCTGCGGAGGCGGTGGCGCGCCCGGCAAGGCGCGTGTG 1080  
Db 341 GlnGlnMetMetThrAlaCysGlnGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGATGAGCCAGGCGCAAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
Db 361 AlaGluAlaMetSerGlnAlaAsnSerValAsnIleMetMetGlnLysSerAsnPheLys 380  
QY 1141 GGGCGCGCGCGCAAGTCAAGTCTTCAACTGCGGCAAGAGGCGCACATGCGCAAGAAC 1200  
Db 381 GlyProArgArgAsnValLysCysPheAsnCysGlyLysGlnGlnHisIleAlaLysAsn 400  
QY 1201 TGCCGCGCGCGCGCAAGAGGCGTGTGAAGTGCAGGAGGCGCAAGGCGCAAGTGAAG 1260  
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlnHisIleGlnMetLys 420  
QY 1261 GACTGCACCGAGCGCCAGGCGCACTTCTGCGCAAGATCTGCGCCAGCGCAAGGCGCGC 1320  
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
QY 1321 CCGGCAACTTCTTGCAGAACCGCAGCGAGCGCGCGCCCGCCACCGTGCCACCGCCCCC 1380  
Db 441 ProGlnAsnPheLeuGlnAsnArgSerGlnProAlaAlaProThrValProThrAlaPro 460  
QY 1381 CCGCGCAGAGCTTCCGCTTGCAGAGAGACCAACCCCGCGCCCGCAAGAGCGCCAAAGAC 1440  
Db 461 ProAlaGlnSerPheArgPheGlnGluThrThrProAlaProLysGlnGlnProLysAsp 480  
QY 1441 CGCGAGCCTTACCGCGAGCGCGCTGACCGCGCTGCGAGCGCTGTTCGAGCGCGCGCTG 1500  
Db 481 ArgGluProTyrArgGluProLeuThrAlaLeuArgSerLeuPheGlySerGlyProLeu 500  
QY 1501 AGCCAG 1506  
Db 501 SerGln 502  
RESULT 2  
AAE37600  
ID AAE37600 standard; protein; 492 AA.  
XX  
AC AAE37600;  
XX  
DT 23-OCT-2003 (revised)  
DT 27-AUG-2003 (first entry)  
XX  
DE HIV-1 subtype C isolate Du422 Gag protein.  
XX  
KW Regulatory gene; accessory gene; HIV; human immunodeficiency virus;  
KW vaccine; infection; gene therapy; Gag.  
OS Human immunodeficiency virus 1.  
XX  
PN WO2003037919-A2.

XX 08-MAY-2003.  
PD  
XX  
XX 31-OCT-2002; 2002WO-IB004550.  
PF  
XX  
PR 31-OCT-2001; 2001ZA-00008978.  
XX  
XX (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.  
PA (UYCA-) UNIV CAPE TOWN.  
PA  
XX  
PI Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;  
XX  
DR WPI; 2003-430497/40.  
DR N-PSDB; AAD29257.  
XX  
XX  
PT New molecules comprising HIV-1 subtype isolate regulatory/accessory  
PT genes, useful for manufacturing a vaccine for treating or preventing HIV  
PT infection.  
XX  
XX  
PS Disclosure; Page 91-92; 97pp; English.  
XX  
XX The invention relates to molecules comprising HIV-1 subtype isolate  
CC regulatory/accessory genes (tat, nef and rev genes) and modifications and  
CC derivatives thereof. The invention also provides proteins encoded by such  
CC genes. Sequences of the invention are useful for manufacturing vaccines  
CC for treating or preventing human immunodeficiency virus (HIV) infections.  
CC They are also useful in gene therapy. The present sequence is HIV-1  
CC subtype C isolate Du422 Gag protein. (Updated on 23-Oct-2003 to  
CC standardise OS field)

XX SQ Sequence 492 AA;

Alignment Scores:

Pred. No.:	2.92e-132	Length:	492
Score:	2464.00	Matches:	466
Percent Similarity:	95.62%	Conservative:	14
Best Local Similarity:	92.83%	Mismatches:	12
Query Match:	86.76%	Indels:	10
DB:	6	Gaps:	2

US-09-475-704A-4 (1-1509) x AAE37600 (1-492)

QY 1 ATGGGCGCGCGCGCATCTCGCGGCGAGAGCTGACAACTGGAGAGATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGCGCAAGACACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysHisIleTyrMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGGGGCTTGCCTGAACCCCGGCTGCTGAAGACCGCGAGGCTGCAAGCAGATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGACGCGCGCGCTGAGACCGGACCGAGAGAGCTGCGCAGCTGTACAAC 240  
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGCGCGGACGAGGTCCGCGACACCAAGAGGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisGlyLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGACAAGATCGAGAGAGAGAGACAGAACTACCCATCGTCAGAACTGACGAGGCGCAGATGGTG 360  
DB 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCGGACGCGCAAGGTGAGCCAGAACTACCCATCGTCAGAACTGACGAGGCGCAGATGGTG 420  
DB 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACGAGGCCATCAGCCCCGCGACACCTGAAGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160

QY 481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCCAGAC 540  
DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGCGGCGCCACGAGCCGCGCATGAGATGCTGAAGAC 600  
DB 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGACCGCTGACACCCCGTGCAGCGCGCCCGTG 660  
DB 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgValHisProValHisAlaGlyProIle 220  
QY 661 GCGCCCGCGCAGATGCGCGACCCCGCGCGCAGCGACATCGCCGCGCCACGACACCTG 720  
DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGACGAGATGCGCTGGATGACCGACCAACCCCGCGTGGCGCGGCGACATCTACAAG 780  
DB 241 GlnGluGlnIleAlaThrMetThrSerAsnProProIleProValGlyAspIleTyrLys 260  
QY 781 CCGTGATCATCTGGGCTGAAACAAGATCGTGGGATGTACAGCCCGGTGAGCATCTG 840  
DB 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGCAGGCGCCCAAGAGCGCTTCGCGCACTACGTGACCGCTTCTTCAAGAC 900  
DB 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 300  
QY 901 CTGCGCGCGGAGAGCGCCACCGCAAGCACTGTAAGACTGATGACCGACACCTGCTGTG 960  
DB 301 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAACGCCAACCCCGCACTGCAAGACCATCTGCGCGCTCTCGGCGCGCGCCACCTG 1020  
DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGCGCGCGGTGGCGCGCGCGCCCAAGGCGCGCTGTG 1080  
DB 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGACCCGCAAGGCCAACAGCGTGAACATCATGATGACAGAAAGCACTTCAAG 1140  
DB 361 AlaGluAlaMetSerGlnAlaAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380  
QY 1141 GCGCCCGCGCGCAAGCTCAAGTCTTCAACTGCGCGCAAGAGGGCCACATCGCCCAAGAC 1200  
DB 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 400  
QY 1201 TGC CGCGCGCGCGCAAGAGGGGCTGCTGAAGTGGCGCAAGAGGGCCACAGATGAAG 1260  
DB 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420  
QY 1261 GACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCAGCCACCAAGGCGCGC 1320  
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTyrProSerHisLysGlyArg 440  
QY 1321 CCGCGCAACTTCTGTCAGAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
DB 441 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454  
QY 1381 CCGCGCGAGAGCTTCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGCG 1440  
DB 455 ProAlaGlnSerPheArgPheGluGluThrThrProAlaProLysGlnGluProIleGlu 474  
QY 1441 CCGGAGCCCTTACCGCGAGCGCGCTGACCGCGCTGCGCGAGCGCTGTTGCGAGCGCGCGCTG 1500  
DB 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
QY 1501 AGCCAG 1506  
DB 491 SerGln 492

RESULT 3  
ADP02848  
ID ADP02848 standard; protein; 492 AA.  
XX  
AC ADP02848;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Wild type HIV-1 gag protein from strain Du499.  
XX  
KW anti-HIV; vaccine; HIV; Gag; prophylaxis; immunogenic response.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO2004050691-A2.  
XX  
PD 17-JUN-2004.  
XX  
PF 04-DEC-2003; 2003WO-IB005634.  
XX  
PR 04-DEC-2002; 2002ZA-00009830.  
XX  
PA (UYCA-) UNIV CAPE TOWN.  
XX (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.  
XX  
PI Jaffray A, Williamson A, Rybicki EP;  
XX  
DR WPI; 2004-450716/42.  
XX  
DR N-PSDB; ADP02846.  
XX  
PT New vector including a nucleotide sequence encoding the Gag polypeptide,  
PT useful as a vaccine for the treatment or prophylaxis of HIV infection in  
a mammal.  
XX  
PS Disclosure; SEQ ID NO 4; 37bp; English.  
XX  
XX The invention relates to a vector including a nucleotide sequence  
CC encoding an HIV Gag polypeptide, where the nucleotide sequence encoding  
CC the Gag polypeptide comprises a sequence having at least 90% sequence  
CC identity to 1549 (S1) or 1479 (S2) bp defined in the specification. The  
CC vector is useful as a vaccine for the treatment or prophylaxis of HIV  
CC infection in a mammal, where the vaccine induces an immunogenic response  
CC to the virus-like particles in a suitable susceptible host. This sequence  
CC corresponds to an HIV-1 gag protein sequence used in the method of the  
invention.  
XX  
SQ Sequence 492 AA;  
XX  
Alignment Scores:  
pred. No.: 3.33e-132 Length: 492  
Score: 2463.00 Matches: 466  
Percent Similarity: 95.42% Conservative: 13  
Best Local Similarity: 92.83% Mismatches: 13  
Query Match: 86.73% Indels: 10  
DB: 8 Gaps: 2  
US-09-475-704A-4 (1-1509) x ADP02848 (1-492)  
QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGGAGAACTGGAAGTGGAGAAATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGluIlyLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGCGCAGAAAGCACTGTAAGCACTGATGTTGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysIlyShiTryMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAGGGCTTGGCCCTGAACCCCGGCTGTGAGACCGCGAGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGACGCGCGCGCTGAGACCGGACCGAGAGAGCTGGCAGCTGTACAAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80

QY 241 ACCGTGGCCACCTGTATGTCGTCGACGCGCGCATCGAGGTCCGCAACAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGACAAGATCGAGGAGGACAGAACAACTCCAGCAGAGAACCCAGCGCCCAAGAG 360  
Db 101 LeuAspLysIleGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGACGGCAGGTGAGCGCAGAACTACCCCATCGTGACAGAACTGACGGCCAGATGTG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACCAGGCCATCAGCCCCCGCACCTGAAGCCCTGGTGAAGGTGATCGAGAGAGGCC 480  
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGGGCGCCACCCCCAGAG 540  
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGCGCGCCACAGCGCCCATGACAGATGCTGAAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGACCGCCTGCACCCCGTGACGCGCGCCCGTG 660  
Db 201 ThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCCCGCGCGCAGATGCGCGACCCCGCGCGCAGACATCGCGCGCGCCAGCACACCTG 720  
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGACGAGATCGCCTGATGACCAAGCAACCCCGCGCGCGTGGCGCGCATCTACAAG 780  
Db 241 GlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCTGGAACAAGATGTCGGATGTACAGCCCGTGCATCTCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGCGCGCGCCCAAGAGCCCTTCGCGCACTACGTGACCGCTTCTCAAGACC 900  
Db 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhehelysThr 300  
QY 901 CTGGCGCGCGCAGCGCCAGCCAGGACGTAAGAACTGATGACCGAGACCTGTGCTG 960  
Db 301 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuVal 320  
QY 961 CAGAAGCGCAACCCCGACTGCAAGACCATCTGCGCGCTGTGCGCGCCGCGCCACCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340  
QY 1021 GAGGAGATGATGACCGCCTGCGCAGGGCGTGGCGCGCCGCGCCCAAGGCGCGTGTG 1080  
Db 341 GlnGluMetMetThrAlaCysGlnGlnGlyValGlyGlyProGlnHisLysAlaArgValLeu 360  
QY 1081 GCCGAGGCGATGAGCCAGGSCAAGCGTGAACATCATGATGCAAGAGCAATTCAAG 1140  
Db 361 AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380  
QY 1141 GGCCCCCGCGCGCAAGTCAAGTGTCTCAACTGCGCGCAAGGAGGCGCCACATGCGCAAGAAC 1200  
Db 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlnHisIleAlaArgAsn 400  
QY 1201 TGCCCGCGCGCGCCAGAGAGGCTGTGAAGTGCGGCAAGAGGCGCCACAGATGAAG 1260  
Db 401 CysArgAlaProArgLysLysLysCysTyrLysCysGlyLysGluGlnHisIleGlnMetLys 420  
QY 1261 GACTGCACCGAGCGCCAGGSCCAACTTCTGGGCAAGATCTGGCCAGCCACAGAGGCGCGC 1320  
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440



QY	1081	GCCGAGGGCATGAGCCAGGCCCAACGCGTGAACATCATGATGAGAAAGCAACTTCAAG	1140
Db	363	AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPhelys	382
QY	1141	GGCCCCCGCGCAACGTCAAGTGTCTTCACTGCGCAAGAGGGCCACATCGCCAAGAAC	1200
Db	383	GlyProArgArgIleValIlysCysPheAsnCyseGlyLysGluGlyHisIleAlaArgAsn	402
QY	1201	TGCGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTGGCGCAAGAGGGCCACCATGAAG	1260
Db	403	CysArgAlaProArgGlyLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys	422
QY	1261	GACTGCACCCGAGCGCCAGGCCCACTTCTGGGCAAGATCTGGCCCCAGCCACAGGGCCGC	1320
Db	423	AspCysThrGlnArgGlnAlaAsnPhelenuGlyLysIleTrpProSerHisLysGlyArg	442
QY	1321	CCCGGCAACTTCTGTCAGAACCCGAGCGAGCCCGCCGCCCAACCGTGCACCGCCCCC	1380
Db	443	ProGlyAsnPhelenuGlnAsnArgProGlu-----ProThrAlaPro	456
QY	1381	CCCGCCGAGAGCTTCCGCTTCGAGGAGAACACCCCCGCCCAAGAGAGCCCAAGGAC	1440
Db	457	ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProIleGlu	476
QY	1441	CGCGAGCCCTACCGCGAGCCCTGACCGCCCTGGCGACGCTGTTCGGACGGGCCCTCG	1500
Db	477	-----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu	492
QY	1501	AGCCAG 1506	
Db	493	SerGln 494	
RESULT 5			
AAE18303			
ID	AAE18303	standard; protein; 492 AA.	
XX			
AC	AAE18303;		
XX			
DT	29-AUG-2003	(revised)	
DT	07-MAY-2002	(first entry)	
XX			
DE	Human immunodeficiency virus type 1 (HIV-1) Clade C gag protein.		
XX			
KW	Human immunodeficiency virus; vaccine; HIV infection; immune response;		
KW	alpha-virus replicon; therapy; gag gene.		
XX			
OS	Human immunodeficiency virus 1.		
XX			
PN	WO200203917-A2.		
XX			
PD	17-JAN-2002.		
XX			
PF	09-JUL-2001; 2001WO-US021701.		
XX			
PR	07-JUL-2000; 2000US-0216995P.		
XX			
PA	(ALPH-) ALPHAVAX INC.		
PA	(UYNC-) UNIV NORTH CAROLINA.		
PI	Olmsted R, Keith P, Dryga S, Caley I, Maughan M, Johnston R;		
PI	Davis N, Swansstrom R;		
XX			
DR	WPI; 2002-171664/22.		
DR	N-PSDB; AAD29133.		
XX			
PT	Composition useful for treating or preventing HIV infections, comprises		
PT	two or more isolated nucleic acids encoding env, gag or pol gene product		
PT	of HIV or immunogenic fragment of the gene products.		
XX			
PS	Example 5; Page 165-166; 201pp; English.		
XX			
CC	The invention relates to a composition comprising isolated nucleic acids		
CC	encoding env, gag or pol gene product of human immunodeficiency virus or		

CC	immunogenic fragment of the gene products. The gag gene product is
CC	modified to inhibit formation of virus-like particles containing gag gene
CC	product and their release from cells, and the pol gene product is
CC	modified to inhibit reverse transcriptase activity. The invention also
CC	relates to a method for producing an alpha-virus replicon particle used
CC	in vaccines. The composition is useful for inducing an immune response to
CC	human immunodeficiency virus (HIV) or for treating or preventing HIV
CC	infection in a subject. The alpha-virus replicon particle is useful in a
CC	vaccine. The composition is useful for administering a protein or peptide
CC	to a subject. A composition comprising heparin affinity-purified alpha-
CC	virus replicon particle is useful as a clinical trial material and as a
CC	commercial product. The present sequence is HIV-1 Clade C gag protein.
CC	(Updated on 29-AUG-2003 to standardise OS field)
XX	
SQ	Sequence 492 AA;
	Alignment Scores:
	Pred. No.:           7.31e-132       Length:           492
	Score:             2457.00          Matches:          465
	Percent Similarity: 95.22%       Conservative:    13
	Best Local Similarity: 92.63%      Mismatches:     14
	Query Match:       86.51%       Indels:           10
DB:	5                                   Gaps:            2
US-09-475-704A-4 (1-1509) x AAEI8303 (1-492)	
QY	1 ATGGCGGCCCGCGGCATCTCGCGCGGAGACTGGACAATTGGAGAAGATCCGC 60
Db	1 MetAlaAlaArgAlaSerIleLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20
QY	61 CTGCGCCCCGCGCGCAAGACACTATCATGTGAAGCACTGTGTGGCCAGCGCGAG 120
Db	21 LeuArgProGlyGlyLysLysHisArgMetLeuYshHisIleValTPrAlaSerArgL 40
QY	121 CTGAGAGGGCTTCGCGCTTGAAACCGCGGCTGTGTGAAGACCGCGGAGGCTGCAAGCAGATC 180
Db	41 LeuGluArgPheAlaLeuAsnProGlyLeuGluThrSerGluGlyCysLysGlnIle 60
QY	181 ATGAAGCAGCTGCAGCCCGCGCTGCAGACC CGGSCAGAGAGCTGCAGCCTGTCAAC 240
Db	61 MetLysGlnLeuGlnProAlaLeuGlnThrgLYThrGluLnuLysSerLeuTYrAsn 80
QY	241 ACCGTGGCCACCCCTGTACTGCGTGCAGCGCGGCATCGAGTCCCGCACACCAAGAGGCC 300
Db	81 ThrValAlaThrLeuTYrCySvalHisGluYsIleGluValArgAspThrLysGluAla 100
QY	301 CTGGACAAGATGAGAGAGAGAGACAAGTCCAGCAGAGAACCCAGAGGCCAAGAGAG 360
Db	101 LeuAspLysIleGluGluGlnAsnLysCySGlnGlnLysThrGlnGlnAlaLysAla 120
QY	361 GCCGACGGCAGAGGTGAGCCAGAACTACCCCATGTGCAGAACTTCGACGGCCAGATGTG 420
Db	121 AlaAspGlyLysValSerGlnAsnTYrProIleValGlnAsnLeuGlnGlyGlnMetVal 140
QY	421 CACCAAGGCATCAGCCCCCGCACCCCTGAAACGCGCTGGTGAAGTGATCGAGGAAAGCCC 480
Db	141 HisGlnAlaIleSerProArgThrLeuAsnAlaTPrValLysValIleGluGluLysAla 160
QY	481 TTCAGCCCCGAGGTGATCCCCCATGTTCACCGCGCTGAGCGAGGGCGCCACCCCCCAGGAC 540
Db	161 PheSerProGluValIlePhePrometPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY	541 CTGAACACAGATTTGAACAACCGTGGGGCGGCACAGCGCCGACATGCATGCTGAAGGAC 600
Db	181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIalaMetGlnMetLeuLysAsp 200
QY	601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGCGCTGCACCCCGTGACAGCGCGCCCGTG 660
Db	201 ThrIleAsnGluGluAlaIalaglutPrAspArgLeuHisProValHisIalaglyProIle 220
QY	661 GCCCCCGGCAGATGCGCGACCCCGCGGACGACATCGCGCGGCACAGCACCCCTG 720
Db	221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaglyThrThrSerThrLeu 240

```

QY 721 CAGAGCAGATCGCTTGATGACCAAGCAACCCCGTGCCTGGGCGATCTACAAG 780
Db 241 GlnGlnGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyr 260
QY 781 CGGTGATCATCTGGGCTGAACAAGATCGTGGATGATACAGCCCGGAGCATCTG 840
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280
QY 841 GACATCCGCCAGGCGCCCAAGAGCCCTTCCGCACTAGTGAGCCGCTTCTTCAAGAC 900
Db 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheTyr 300
QY 901 CTGCGCGCCGAGCAGGCAAGCCCAAGACGTAAGTGAAGTGAAGTGAAGTGAAGTGA 960
Db 301 LeuArgAlaGlnGlnIleAlaThrGlnGlnValLysAsnTrpMetThrAspThrLeuLeuVal 320
QY 961 CAGAACGCCCAACCCCGCACTGCAAGACCACTCTGCGGCTTCCGCGCCCGGCGCACCTG 1020
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340
QY 1021 GAGGAGATGATGACCGCTGCCAGGCGCGTGGGCGCGCCCGCCCAAGCCCGCTGCTG 1080
Db 341 GluGluMetMetThrAlaCysGlnGlnValGlyGlyProGlyHisLysAlaArgValLeu 360
QY 1081 GCCGAGCGCATGAGCCAGCCCAACAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140
Db 361 AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPhelys 380
QY 1141 GGGCCCCGGCGCAACGTCAGTGTCTTCACTGCGGCAAGAGGCGCCACATCGCCAAAGAAC 1200
Db 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGlnGlnIleAlaArgAsn 400
QY 1201 TGCCCGCGCCCCCGCAAGAGGGTGTGCTGGAAGTCCGCAAGAGGCGCCACCATGTAAG 1260
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlnIleMetLys 420
QY 1261 GACTGCAACCGAGCGCGCCCAACTTCTTGGGCAAGATCTGGCCAGCCCAAGAGGCGCGC 1320
Db 421 AspCysThrGlnArgGlnAlaAsnPhenLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGCGCAACTTCTCTGCAAGAACCGCAGCGAGCCCGCGCCCGCCACCGTGCACCGCCGCC 1380
Db 441 ProGlyAsnPhenLeuGlnAsnArgProGlu-----ProThrAlaPro 454
QY 1381 CCGCGCGAGAGCTTCCGCTTGCAGAGAACCAACCCCGCGCCCAAGCAGAGCCCAAGGAC 1440
Db 455 ProAlaGluSerPheArgPheGlnGlnThrThrProAlaProLysGlnGlnProIleGln 474
QY 1441 CGCGAGCCCTTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTGGGAGCGGCGCCCTG 1500
Db 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490
QY 1501 AGCCAG 1506
Db 491 SerGln 492

```

```

RESULT 6
AAM48948
ID AAM48948 standard; protein; 635 AA.
XX
AC AAM48948;
XX
DT 29-AUG-2003 (revised)
DT 19-APR-2002 (first entry)
XX
DE HIV-1 subtype C isolate Du422 gag protein.
XX
KM HIV-1 subtype C; vaccine; HIV infection; AIDS; gag; antiviral.
XX
OS Human immunodeficiency virus; type I.
XX
FH Key Location/Qualifiers

```

```

FT Misc-difference 523
FT /note= "encoded by TAA"
FT Misc-difference 526
FT /note= "encoded by TAG"
FT Misc-difference 557
FT /note= "encoded by TAA"
FT Misc-difference 612
FT /note= "encoded by TGA"
XX
PN WO200204494-A2.
XX
PD 17-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-IB001208.
XX
PR 07-JUL-2000; 2000US-0216995P.
PR 10-JUL-2000; 2000ZA-00003437.
PR 15-SEP-2000; 2000ZA-00004924.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
PA (UYCA-) UNIV CAPE TOWN.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Williamson C, Swanstrom RI, Morris L, Karim SA, Johnston RE;
XX
DR WPI; 2002-171700/22.
DR N-PSDB; AAL41591.
XX
PT Selecting HIV-1 subtype C isolates, which are useful in developing
PT vaccines against HIV infection, comprises isolating viruses with high
PT sequence identity to a consensus sequence whose phenotype is associated
PT with the HIV subtype.
XX
PS Claim 17; Page 64-66; 69pp; English.
XX
CC The present invention relates to a process for the selection of human
CC immunodeficiency virus (HIV) subtype isolates for use in the development
CC of a prophylactic and/or therapeutic pharmaceutical composition. The
CC process involves selecting isolated virus or viruses with a high sequence
CC identity to a consensus sequence and a phenotype which is associated with
CC transmission for the particular HIV subtype. The composition can be used
CC in the production of vaccines against HIV. The present sequence is the
CC HIV-1 subtype C isolate Du422 gag protein. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
SQ Sequence 635 AA;

```

```

Alignment Scores:
Pred. No.: 7.45e-132 Length: 635
Score: 2457.00 Matches: 465
Percent Similarity: 95.22% Conservative: 13
Best Local Similarity: 92.63% Mismatches: 14
Query Match: 86.51% Indels: 10
DB: 5 Gaps: 2

```

US-09-475-704A-4 (1-1509) x AAM48948 (1-635)

```

QY 1 ATGGGCGCCCGCGCAGCATCTTGGCGGCGGAGAGCTGAGCAAGTGGAGAGATCCGC 60
Db 31 MetAlaAlaArgAlaSerIleLeuArgGlyGlnLysLeuAspLysTrpGlnLysIleArg 50
QY 61 CTGGCGCCCGGCGCAAGAGCACTAGTGAAGCACTGTGTGGGCGGAGCGCGAG 120
Db 51 LeuArgProGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 70
QY 121 CTGGAGGGCTTGGCTTGAACCCCGGCTGCTGAGAGCCGCGGAGGCTGCAAGCATC 180
Db 71 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnGlyCysLysGlnIle 90
QY 181 ATGAAGCAGCTGAGCGCCCGCTGCAAGACCGGCAACCGAGAGCTGCGAGCTGTACAAC 240
Db 91 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuLysSerLeuTyrAsn 110

```

QY 241 ACCGTGGCCACCTGTACTGCGTGACAGCCGGCATCGAGTCCGGACACCAAGAGAGCC 300  
Db 111 ThrValAlaThrLeuTyrcysValHisGlnLysIleGluValArgAspThrLysGluAla 130  
QY 301 CTGGACAAGATCGAGAGAGAGACAGAACAAAGTCCAGCAGAGAAGACCAGAGGCCAAGAG 360  
Db 131 LeuAspLysIleGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 150  
QY 361 GCCGACGGCAGAGGTGAGCCAGAACTACCCCATCGTGACAGACCTGACAGGCCAGATGTTG 420  
Db 151 AlaAspGlyLysValSerGlnAsnTyrcProIleValGlnAsnLeuGlnGlyGlnMetVal 170  
QY 421 CACCAAGCCATCAGCCCCCGCAGACCCCTGAACGCTGGTGAAGTGATCGAGAGAAGAGCC 480  
Db 171 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 190  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTTACCGCCCTGAGCGAGGGCGCCACCCCGCAGGAC 540  
Db 191 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 210  
QY 541 CTGAACACGATGTTGAACACCGTGGCGCGCCACAGGCCCATGACAGATGCTGAAGGAC 600  
Db 211 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLysAsp 230  
QY 601 ACCATCAACGAGAGAGCCCGCAGTGGGACCGCTGACCCCGTGACAGCCGCGCCGTG 660  
Db 231 ThrIleAsnGluGluAlaIleGluTrpAspArgLysHisProValHisAlaGlyProIle 250  
QY 661 GCCCGCGCGCAGATGCGCGGACCCCGCGCGCAGCGACATGCGCGCGCCACAGACCCCTG 720  
Db 251 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 270  
QY 721 CAGAGACGATCGCTGATGACCAAGACCCCGCTGCGCGCGCAGCATCTTCAAG 780  
Db 271 GlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyrls 290  
QY 781 CGGTGGATCATCTGGGCTTGAACAAGATCGTGCGGATGTACAGCCCGTGAGCATCTCTG 840  
Db 291 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrcSerProValSerIleLeu 310  
QY 841 GACATCCGCCAGGCGCCCAAGAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAGACC 900  
Db 311 AspIleArgGlnGlyProLysGluProPheArgAspTyValAspArgPhePheLysThr 330  
QY 901 CTGCGCGCGCGAGCAGGCGCAACCCAGACGTGAAGATCGATGACCGACCCCTGTGTG 960  
Db 331 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuVal 350  
QY 961 CAGACGCCCAACCCCGACTGCAAGACATCTCGCGGCTCTCGGCGCGCGCGCCACCCCTG 1020  
Db 351 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 370  
QY 1021 GAGGAGATGATGACCGCTTCCAGAGCGGTGGCGCGCCCGCCACAGAGCGCGGTGTG 1080  
Db 371 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 390  
QY 1081 GCCGAGGCGGATGAGCCAGGCGCAACCGCTGAACATCATGATGACAGAGAAGCACTTCAAG 1140  
Db 391 AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 410  
QY 1141 GGGCCCCGGCGCAAGTCAAGTGTCTTCACTGCGGCAAGAGGGGCCACATCGGCAAGAAC 1200  
Db 411 GlyProArgArgIleValLysCysPheAsnCyseGlyLysGluGlyHisIleAlaArgAsn 430  
QY 1201 TGGCGCGCGCGCGCAAGAGGGGTGTGGAAGTGGCGCAAGAGGGGCCACCAAGTGAAG 1260  
Db 431 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleGlnMetLys 450  
QY 1261 GACTGCACCGAGCGCCAGGCCCAACTTCTGTGGGCAAGATCTGGGCCACCAAGGGCGCGC 1320  
Db 451 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 470  
QY 1321 CCGCGCAACTTCTGTGAGAACCGCAGAGCGCGCGCGCCCGCCACCGTGTGCCACCGCGCCC 1380

Db 471 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 484  
QY 1381 CCGCGCAGAGCTTCCGCTTGAAGAGACCAACCCCGCCCGCCAGCAGAGGCCCAAGAGAC 1440  
Db 485 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProIleGlu 504  
QY 1441 CGCGAGCCCTACCGCGAGCCCGCTGACCGCCCTGCGCGCAGCCCTGTTGCGAGCGGCCCTG 1500  
Db 505 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 520  
QY 1501 AGCCAG 1506  
Db 521 SerGln 522  
RESULT 7  
ADCT2879  
ID ADCT2879 standard; protein; 492 AA.  
XX  
AC ADCT2879;  
DT 18-DEC-2003 (first entry)  
XX  
DE HIV-1 gag protein containing an Ncp sequence, SEQ ID NO 109.  
XX  
KW complex; HIV nucleocapsid protein 7; Ncp7; HIV-psi-site; anti-HIV;  
KW human immunodeficiency virus; HIV.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO2003060098-A2.  
XX  
PD 24-JUL-2003.  
XX  
PF 10-JAN-2003; 2003WO-US000801.  
XX  
PR 11-JAN-2002; 2002US-0347369P.  
XX  
PA (ACHI-) ACHILLION PHARM INC.  
XX  
PI Beuchter D, Hou X, Marlor CW, Rice WG, Yang W;  
XX  
DR WPI; 2003-646042/61.  
XX  
PT Method of determining whether a compound inhibits formation of complex  
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
PT oligonucleotide by comparing amount of complex formed in presence/absence  
PT of compound.  
XX  
PS Claim 2; SEQ ID NO 109; 105bp; English.  
XX  
CC The invention relates to a novel method of determining whether a compound  
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (Ncp7)  
CC polypeptide and an HIV-psi-site oligonucleotide. The method involves  
CC adding the HIV-psi-site oligo to a mixture of the Ncp7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an Ncp sequence of the invention.  
XX  
SQ Sequence 492 AA;  
Alignment Scores:  
Pred. No.: 6.79e-131 Length: 492  
Score: 2440.00 Matches: 464  
Percent Similarity: 94.42% Conservative: 10  
Best Local Similarity: 92.43% Mismatches: 18  
Query Match: 85.92% Indels: 10  
DB: 7 Gaps: 2

US-09-475-704A-4 (1-1509) x ADC72879 (1-492)

QY 1 ATGGGCGCCGCGCGCAGCATCTGCGCGCGAGAAAGCTGACAAGTGGAGAAAGATCCGC 60  
 |||||  
 Db 1 MetGlyAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
 QY 61 CTGCGCGCCGCGCGCAAGACACTACATGCTGAAGCACTGTGTGGCCAGCGCGAG 120  
 |||||  
 Db 21 LeuArgProGlyGlyLysLysLysTrpArgLeuLysHisLeuValTrpAlaSerArgGlu 40  
 QY 121 CTGAGGGCTTTCGCTGAAACCCGCGCTGTGAGACCGCGGAGGCTGCAAGCAGATC 180  
 |||||  
 Db 41 LeuGluArgPheAlaLeuAsnSerGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60  
 QY 181 ATGAGCAGCTGCAGCGCGCCCTGCAGACCGGACGAGAGCTGCGCAGCTGTACAC 240  
 : : : : :  
 Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTrpAsn 80  
 QY 241 ACCGTGGCCACCTCTTACTGCGTGCACGCGCGGATCGAGGTCCGCGACACCAAGAGGCC 300  
 |||||  
 Db 81 ThrValAlaThrLeuTrpCysValHisAlaGlyIleGluValArgAspThrLysGluAla 100  
 QY 301 CTGACACAAGATCGAGAGGAGGAGACAAGTCCACAGAGAAGACCCAGAGGCCAAGAG 360  
 |||||  
 Db 101 LeuAspLysIleGluGlnGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysGlu 120  
 QY 361 GCGGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGTG 420  
 |||||  
 Db 121 AlaAspGlyLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
 QY 421 CACCAAGGCCATCAGCCCCCGCACCCCTGAACGCGCTGGTGAAGGTGATGAGAGAGGCC 480  
 |||||  
 Db 141 HisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
 QY 481 TTCAGCCCCGAGGTGATTCCTCATGTTCAACGCGCTGAGCGAGGGCGGCCACCCCGCAGAC 540  
 |||||  
 Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
 QY 541 CTGAACAACGATGTTGAACACCGTGGGGGGCCACACGAGCGCCATGCAATGCTGAAGAC 600  
 |||||  
 Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
 QY 601 ACCATCAACGAGGAGGCGCGCGAGTGGGACCGCCTGCACCCCGTGCAGGCGCGGCCGCTG 660  
 |||||  
 Db 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProVal 220  
 QY 661 GCGCGCGCGCAGATGCGCGACACCCCGCGCGCAGCGACATCGCGGCGGCCACACGACCTG 720  
 |||||  
 Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
 QY 721 CAGAGCAGATCGCTGGATGACAGCAACCCCGCTGCGCGTGGGCGACATTCACAAG 780  
 |||||  
 Db 241 GlnGluGlnIleThrTrpMetThrSerAsnProProValProValGlyAspIleTrpLys 260  
 QY 781 CCGTGGATCATCTTGGGCTTGAACAAGATCGTGGATGTACAGCCCGTGAGCATCTG 840  
 |||||  
 Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTrpSerProValSerIleLeu 280  
 QY 841 GACATCCGCGCAGGCGCCCAAGAGACCTTCGCGCATACGTGACCGCTTCTTCAAGAC 900  
 |||||  
 Db 281 AspIleLysGlnGlyProLysGluProPheArgAspTrpValAspArgPhePheLysVal 300  
 QY 901 CTGCGCGCGCAGAGCAGGCCACCCAGAGAGTGAAGAACTGATGACCGAGACCTGTGCTG 960  
 |||||  
 Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 320  
 QY 961 CAGAACGCCAACCCGCACTGCAAGACATCTGCGCTCTCGCGCCCGCGCGCACCTG 1020  
 |||||  
 Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProAlaAlaSerLeu 340  
 QY 1021 GAGGAGATGATGACCGCTGCCAGGGCGTGGCGCGCGCGCCACAGAGCCCGCGCTGCTG 1080  
 |||||

Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgValLeu 360  
 QY 1081 GCCGAGCGGATGAGCCAGGCCACAGCGGTGAACATCATGATGCAGAAAGCAACTTCAAG 1140  
 |||||  
 Db 361 AlaGluAlaMetSerGlnAlaAsnThrThrAsnIleMetGlnLysSerAsnPheLys 380  
 QY 1141 GCGCGCGCGCGCAAGCTCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATCGCCACAGAC 1200  
 |||||  
 Db 381 GlyProArgArgThrValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLysAsn 400  
 QY 1201 TGCGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGGCGGCAAGAGGGCCACAGATGAAG 1260  
 |||||  
 Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420  
 QY 1261 GACTGCACCGAGCGCCAGGCCAACTTCTGCGCAAGATCTGCGCCACCAAGGCGCGC 1320  
 |||||  
 Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
 QY 1321 CCGGCAACTTCTGCAAGAACCGCAGCGAGCGCGCGCGCCACCGTGCACCGCGCGC 1380  
 |||||  
 Db 441 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 454  
 QY 1381 CCGCGCGAGAGCTTCCGCTTGCAGAGAGACCAACCCCGCGCCCAAGCAGAGGCCAAGAC 1440  
 |||||  
 Db 455 ProAlaGluSerPheArgPheGluGlnIleThrThrProAlaProLysGlnGluProLysAsp 474  
 QY 1441 CGCGAGCCCTACCGGAGCGCGCGCTGACCGCGCGCTGCGCAGCCTGTTCGCGAGCGCGCCTG 1500  
 |||||  
 Db 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
 QY 1501 AGCCAG 1506  
 |||||  
 Db 491 SerGln 492

## RESULT 8

ADC72880  
 ID ADC72880 standard; protein; 492 AA.  
 XX  
 AC ADC72880;  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE HIV-1 gag protein containing an NCP sequence, SEQ ID No 110.  
 XX  
 KW complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;  
 KW human immunodeficiency virus; HIV.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 PN WC2003060098-A2.  
 XX  
 PD 24-JUL-2003.  
 XX  
 PF 10-JAN-2003; 2003WO-US000801.  
 XX  
 PR 11-JAN-2002; 2002US-0347369P.  
 XX  
 PA (ACHI-) ACHILLION PHARM INC.  
 XX  
 PI Beutcher D, Hou X, Marlor CW, Rice WG, Yang W;  
 XX  
 DR WPI; 2003-646042/61.  
 XX  
 PT Method of determining whether a compound inhibits formation of complex  
 PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
 PT oligonucleotide by comparing amount of complex formed in presence/absence  
 PT of compound.  
 XX  
 PS Claim 2; SEQ ID NO 110; 105pp; English.  
 XX  
 CC The invention relates to a novel method of determining whether a compound  
 CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)  
 CC polypeptide and an HIV-psi-site oligonucleotide. The method involves

CC adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an NCP sequence of the invention.

XX  
SQ Sequence 492 AA;

Alignment Scores:

Pred. No.:	2.52e-130	Length:	492
Score:	2430.00	Matches:	458
Percent Similarity:	94.82%	Conservative:	18
Best Local Similarity:	91.24%	Mismatches:	16
Query Match:	85.56%	Indels:	10
DB:	7	Gaps:	2

US-09-475-704A-4 (1-1509) x ADC72880 (1-492)

```
QY 1 ATGGGCGCCGCGCGCATCTCGCGCGGAGAGAGCTGGACAAGTGGAGAGATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspTrpTrpGluLysIleArg 20
QY 61 CTGGCGCCGCGCGGAGAGAGACTACATGCTGAAAGCACTGTGTGGGCGGAGCGGAG 120
Db 21 LeuArgProGlyGlyLysLysCysTyrMetLeuLysHisIleValTrpAlaSerArgGlu 40
QY 121 CTGGAGGGCTTCGCGCTGAACCCCGGCTGTGGAGACCGCGGAGGCTGCAAGAGATC 180
Db 41 LeuGluArgPheSerLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
QY 181 ATGAAGCAGCTGCACCCCGCGCTGACAGACCGGACCGAGAGAGCTGGCGCAGCTGTACAAC 240
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLeuLysSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGCGCGCATCGAGTCCGCGACACCAAGAGAGCC 300
Db 81 ThrValAlaThrLeuPheCysValHisGluLysIleAlaValArgAspThrLysGluAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGAGACAAGTCCACAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120
QY 361 GCCGAGCGGCAAGGTGAGCGCAAGACTACCCCATCGTGCAGAGAGAGAGAGAGAGAGAG 420
Db 121 AlaAspGlyThrValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMetVal 140
QY 421 CACCAAGGCATCAGCCCCCGGACCCCTGAACGCGCTGGGTGAAGGTGATCGAGAGAGAGCC 480
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCGCTGAGCGAGGCGGCGCAACCCCGAGGAC 540
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY 541 CTGAACAACGATTTGAACACCGGTGGGGCGGCGCACAGGCGGCGCATGCAAGTGTGAAGGAC 600
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGACCGCTGCACCCCGTGCAGGCGGCGCGCTG 660
Db 201 ThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisAlaGlyProIle 220
QY 661 GCCCGCGGCGAGATGCGCGGAGCGCGCGGCGGAGAGACATGCGCGGCGGCGGCGGAGAG 720
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240
QY 721 CAGGAGCAGATCGCTGTGATGACGAGCAACCCCGCGTGCCTGGGCGGAGCATCTACAAG 780
Db 241 GlnGluGlnIleAlaTrpMetThrAsnAsnProValProValGlyAspIleTyrLys 260
```

```
QY 781 CGGTGATCATCTGGCGCTGAACAAGATCGTCCGATGTACAGCCCGGTGACATCCTG 840
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280
QY 841 GACATCCGCGAGGCGCCCAAGAGCGCTTCCCGCACTACCGTGAACCGCTTCTTCAAGACC 900
Db 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPheLysThr 300
QY 901 CTGCGCGCGCGAGAGCGCAACCCAGACGCTGAAGAACTGATGACCGAGACCTGTGCTG 960
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCTTCCGCGCGCGCGCGCACCTG 1020
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340
QY 1021 GAGGAGATGATGACCGCGCTGCGGAGGCGGTGGCGCGCGCGCGCACAGCGCGCTGCTG 1080
Db 341 GluGluMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgValLeu 360
QY 1081 GCGGAGCGCATGAGCCAGCGCAAGACCATCTGATGATGACGAGAGCAACTTCAAG 1140
Db 361 AlaGluAlaMetSerGlnThrAsnAsnAlaAsnIleMetMetGlnArgSerAsnPheLys 380
QY 1141 GCGCGCGCGCGCAACGTCAAGTGTCTCACTGCGGCAAGAGAGGCGCACATCGCCCAAGAAC 1200
Db 381 GlyProArgArgIleIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArgAsn 400
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGCAGGAGAGAGGCGCACCATGTAAG 1260
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420
QY 1261 GACTGCACCGAGCGCGCAAGCTTCTCGGCGCAAGATCGGCGCGCACCAAGAGGCGCGC 1320
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGCGCACTTCTGCAAGAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 441 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454
QY 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGCG 1440
Db 455 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProArgGlu 474
QY 1441 CGCGAGCGCTACCGCGGAGCGCGCGCTGACCGCGCTGCGGACGCTTTCGGGAGCGGCGCGCTG 1500
Db 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490
QY 1501 AGCCAG 1506
Db 491 SerGln 492
RESULT 9
ADC72910
ID ADC72910 standard; protein; 491 AA.
XX
AC ADC72910;
DT 18-DEC-2003 (first entry)
XX
DE HIV-1 gag protein containing an NCP sequence, SEQ ID No 140.
XX
KW complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;
KW human immunodeficiency virus; HIV.
XX
OS Human immunodeficiency virus 1.
XX
PN WO2003060098-A2.
XX
PD 24-JUL-2003.
XX
PF 10-JAN-2003; 2003WO-US000801.
```

XX 11-JAN-2002; 2002US-0347369P.  
PR (ACHI-) ACHILLION PHARM INC.  
XX Beutler D, Hou X, Marlor CW, Rice WG, Yang W;  
XX WPI; 2003-646042/61.  
XX  
PT Method of determining whether a compound inhibits formation of complex  
PT between HIV nucleocapsid protein 7 polypeptide and HIV psl-site  
PT oligonucleotide by comparing amount of complex formed in presence/absence  
PT of compound.  
XX  
PS Claim 2; SEQ ID NO 140; 105bp; English.  
XX  
CC The invention relates to a novel method of determining whether a compound  
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)  
CC polypeptide and an HIV-psl-site oligonucleotide. The method involves  
CC adding the HIV-psl-site oligo to a mixture of the NCP7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an NCP sequence of the invention.  
XX  
SQ Sequence 491 AA;

Alignment Scores:  
Pred. No.: 7.68e-130 Length: 491  
Score: 2421.50 Matches: 461  
Percent Similarity: 94.82% Conservative: 15  
Best Local Similarity: 91.83% Mismatches: 15  
Query Match: 85.26% Indels: 11  
DB: 7 Gaps: 3

US-09-475-704A-4 (1-1509) x ADCT2910 (1-491)

QY 1 ATGGGGCCCGCCGCGCAGATCTGCGCGCGAGAGCTGAGCAAGTGGGAGAGATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGGCAAGAGCACTACTGTCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGGGCTTTCGCTCTGAACCCCGGCTGTCTGAGACCGCGCGAGGCTGCAAGCAGATC 180  
DB 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGAGCGCGCGCTGTGAGACCGGCAAGAGCTGCGAGCTGTATCAAC 240  
DB 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTCTGTACTGCGTGCAGCGCGGCGATCGAGTCCGCGACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisAlaGluIleGluValArgAspThrLysGluAla 100  
QY 301 CTGACAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 101 LeuAspArgIleGluGluGluGlnAsnLysSerGlnGlnLysThrGlnGlnAlaAsnGlu 120  
QY 361 GCCGAGCGCAAGGTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
DB 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACCAAGCCATCAGCCCGCGCACCCCTGTAAGCGCTGGTGAAGGTGATGAGAGAGAGCC 480  
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160

QY 481 TTCAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGGCCACCCCGAGAGC 540  
DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGCGCGCCACGAGCGCCCATGTCAGATGCTGAAGAC 600  
DB 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGACCCCGCTGAGCGCGCGCCCTG 660  
DB 201 ThrIleAsnGluGluAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCCCGCGCGAGATGCGCGAGCCCGCGCGCGAGCGAGATCGCCGCGCGCCACGACCCCTG 720  
DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerSerLeu 240  
QY 721 CAGAGCAGATCGCTGGATGACCAAGCAACCCCGCTGCGCGCGCGAGCATCTTCAAG 780  
DB 241 GlnGluGlnIleAlaTrpMetThrGlyAsnProProValProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTTGGCGCTGAACAAGATGTCGCGATGTACAGCCCGTGAAGCATCTG 840  
DB 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGAGGGCGCCCAAGAGAGCCCTTCGCGACTGACGTCGACCGCTTCTTCAAGAC 900  
DB 281 AspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLeuThr 300  
QY 901 CTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320  
QY 961 CAGAGCGCAACCCCGCATGCAAGCACTCTGCGCGCTTCTGCGCGCGCGCGCACCTG 1020  
DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 GAGAGATGATGACCGCTGCGAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCTG 1080  
DB 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGATGAGCGAGCGCAAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
DB 361 AlaGluAlaMetSerGlnThrAsnSer---ThrIleLeuMetGlnArgSerAsnPheLys 379  
QY 1141 GGCCCCGGCGCAAGTCAAGTGTCTCACTGCGCGCAAGAGGGCCCATGCGCAAGAC 1200  
DB 380 GlyProLysArgIleValLysCysPheAsnCyseGlyLysGluGlyHisIleAlaLysAsn 399  
QY 1201 TGCGCGCGCGCGCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGGCCACAGATGAG 1260  
DB 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleMetLys 419  
QY 1261 GACTGACCGAGCGCGCGCAACTTCTTGGGCAAGATCTGGCCCGCACCAAGGGCGCC 1320  
DB 420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCGCGCACTTCTTGAAGAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1380  
DB 440 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 453  
QY 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 454 ProAlaGluSerPheArgPheGluGluThrThrProAlaLeuGlnGlnGlyProLysAsp 473  
QY 1441 CGCGAGCCCTACCGGAGCGCGCTGACCGCGCGCTGCGAGCTGTTCGAGCGCGCGCTG 1500  
DB 474 -----ArgGluProLeuThrSerLeuArgSerLeuPheGlySerAspProLeu 489  
QY 1501 AGCCAG 1506  
DB 490 SerGln 491

RESULT 10

ADN36406  
ID ADN36406 standard; protein; 1457 AA.  
XX  
AC ADN36406;  
XX  
DT 15-JUL-2004 (first entry)  
XX  
DE Human protein for anti-HIV vaccine.  
XX  
KW anti-HIV; vaccine; HIV; promoter; viral particle; immunization.  
XX  
OS Homo sapiens.  
XX  
PN WO2004035006-A2.  
XX  
PD 29-APR-2004.  
XX  
PF 17-OCT-2003; 2003WO-US033112.  
XX  
PR 18-OCT-2002; 2002US-0419465P.  
XX  
PA (AARO-) AARON DIAMOND AIDS RES CENT.  
XX  
PI Huang Y, Ho DD, Chen Z;  
XX  
XX WPI; 2004-348328/32.  
DR N-PSDB; ADN36405.  
XX  
XX  
PT Nucleic acid vector comprising at least one HIV sequence operably linked  
PT to a promoter and encoding a protein that does not assemble into viral  
PT particles, useful in immunizing a subject against HIV infection.  
XX  
PS Disclosure; SEQ ID NO 20; 166pp; English.  
XX  
CC The invention relates to a nucleic acid vector comprising at least one  
CC HIV sequence operably linked to a promoter and encoding a protein that  
CC does not assemble into viral particles. The nucleic acid vector is useful  
CC in immunizing a subject against HIV infection. This sequence corresponds  
CC to a peptide used in the invention.  
XX  
SQ Sequence 1457 AA;  
  
Alignment Scores:  
Pred. No.: 1.83e-129 Length: 1457  
Score: 2415.50 Matches: 458  
Percent Similarity: 94.82% Conservative: 18  
Best Local Similarity: 91.24% Mismatches: 15  
Query Match: 85.05% Indels: 11  
DB: 8 Gaps: 3  
  
US-09-475-704A-4 (1-1509) x ADN36406 (1-1457)  
QY 1 ATGGGCGCGCGCGCCAGCATCCGCGCGCGCGAGAGCTGGACAAGTGGAGAGATCCGC 60  
Db |||||  
Db 24 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGlnLysIleArg 43  
QY 61 CTGGCGCGCGCGCGCAAGAAGCACTACATGCTGAAGCACTGTGTGGCCAGCGCGAG 120  
Db |||||  
Db 44 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisIleValTrpAlaSerArgGln 63  
QY 121 CTGGAAGGCTTCGCGCCCTGAACCCCGCGCTGTGAAGACCGCGCAAGGGCTGCAAGCATC 180  
Db |||||  
Db 64 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnGlyCysLysGlnIle 83  
QY 181 ATGAAGCAGCTGAGACCGCGCTGCAAGACCGCGACCGAGAGCTGCGCAGCCTGTACAAC 240  
Db ::::|  
Db 84 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuArgSerLeuPheAsn 103  
QY 241 ACCGTGGCCACCTCTGTACTGCGGTGCACGCGCGCATGAGGTCGCGACCAAGAGGCC 300  
Db |||||  
Db 104 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGlnValArgAspThrLysGlnVal 123  
QY 301 CTGGAACAAGATCGAGGAGGAGCAACAAGTCCAGCAAGAACCCAGAGGCCAAGGAG 360

Db |||||  
Db 124 LeuAspLysIleGlnGlnGlnAlaAsnLysMetGlnGlnLysThrGlnGlnAlaLysGln 143  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATGTGCAAGACCTGCAGGGCCAGATGGTG 420  
Db |||||  
Db 144 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 163  
QY 421 CACCAGGCCATCAGCCCCCGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
Db |||||  
Db 164 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLysAla 183  
QY 481 TTCAGCCCCGAGGTGATCCCAATGTTCAACGCGCTGAGCGAGGGCGCCACCCCGCAGAG 540  
Db |||||  
Db 184 PheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 203  
QY 541 CTGAACACGATGTGAACACCGGTGGCGGCCACAGGGCCCATGCAATGCTGAAGGAC 600  
Db |||||  
Db 204 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 223  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGTGCAGGCGCGCCCGTG 660  
Db |||||  
Db 224 ThrIleAsnGlnGlnAlaIleAlaGlnTrpAspArgValHisProValHisAlaGlyProIle 243  
QY 661 GCGCGCGCGCGCATGCGCGCAACCCCGCGCGAGCATCGCGCGCGCGCCACGACCCCTG 720  
Db |||||  
Db 244 AlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGlyThrThrSerThrLeu 263  
QY 721 CAGGAGCAGATCGCCTGATGACCAACACCCCGCGCGCGCGCGCGCGCATCTACAAG 780  
Db |||||  
Db 264 GlnGlyGlnIleAlaTrpMetThrSerAsnProProValProValGlyGlnIleTyrLys 283  
QY 781 CGGTGATCATCTGGGCGCTGAACAAGATGTCGGATGTACAGCCCGCGAGCATCTCTG 840  
Db |||||  
Db 284 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 303  
QY 841 GACATCCGCGAGGGCGCCAGAGAGCCCTTCGCGCATGAGTGAACCGCTTCTCAAGACC 900  
Db |||||  
Db 304 AspIleLysGlnGlyProLysGlnProPheArgAspTyrValAlaAspArgPhePheLysThr 323  
QY 901 CTGCGCGCGCGAGCGCAAGCCACGAGAGCTGATGATGACCGAGACCTGTGCTGTG 960  
Db |||||  
Db 324 LeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 343  
QY 961 CAGAACGCCAACCCCGCATGCAAGACCATCTCGCGCTTCGGCGCGCGCGCGCCACCTG 1020  
Db |||||  
Db 344 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 363  
QY 1021 GAGGAGATGATGACCGCGCTGCCAGGCGGTGGCGCGCGCGCCACAGGCGCGCTGTG 1080  
Db |||||  
Db 364 GlnGlnMetMetThrAlaCysGlnGlyValGlyLysProAsnHisLysAlaArgValLeu 383  
QY 1081 GCCGAGCGGATGAGCCAGCGCAACAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140  
Db |||||  
Db 384 AlaGlnAlaMetSerGlnAlaAsnGly---ThrIleLeuMetGlnArgSerAsnPheLys 402  
QY 1141 GGCGCGCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCAGAAC 1200  
Db |||||  
Db 403 GlySerLysArgIleValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaArgAsn 422  
QY 1201 TGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACAGATGAAG 1260  
Db |||||  
Db 423 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisGlnMetLys 442  
QY 1261 GACTGACCGAGCGCGCGCAACTTCTGTGGCAAGATCTGGCCCAAGAGGGCGCGC 1320  
Db |||||  
Db 443 AspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProPheHisLysGlyArg 462  
QY 1321 CCGGCAACTTCTGTGGAACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1380  
Db |||||  
Db 463 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 476  
QY 1381 CCGCGCGAGAGCTTCGCTTGGAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGCGC 1440  
Db |||||

Db 477 PROAlaGluSerPheGlyPheGluGluThrThrProAlaProLysGlnGluProLysAsp 496  
QY 1441 CGCGAGCCCTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTGGCAGCGGCCCTTG 1500  
Db 497 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 512  
QY 1501 AGCCAG 1506  
Db 513 SerGln 514  
RESULT 11  
ADC72913  
ID ADC72913 standard; protein; 491 AA.  
XX  
AC ADC72913;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE HIV-1 gag protein containing an Ncp sequence, SEQ ID NO 143.  
XX  
KM complex; HIV nucleocapsid protein 7; Ncp7; HIV-psi-site; anti-HIV;  
KM human immunodeficiency virus; HIV.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO2003060098-A2.  
XX  
PD 24-JUL-2003.  
XX  
PF 10-JAN-2003; 2003WO-US000801.  
XX  
PR 11-JAN-2002; 2002US-0347369P.  
XX  
PA (ACHI-) ACHILLION PHARM INC.  
XX  
PI Beuchter D; Hou X, Marlor CW, Rice WG, Yang W;  
XX  
DR WPI; 2003-646042/61.  
XX  
PT Method of determining whether a compound inhibits formation of complex  
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
PT oligonucleotide by comparing amount of complex formed in presence/absence  
PT of compound.  
XX  
PS Claim 2; SEQ ID NO 143; 105pp; English.  
XX  
CC The invention relates to a novel method of determining whether a compound  
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (Ncp7)  
CC polypeptide and an HIV-psi-site oligonucleotide. The method involves  
CC adding the HIV-psi-site oligo to a mixture of the Ncp7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an Ncp sequence of the invention.  
XX  
SQ Sequence 491 AA;  
Alignment Scores:  
Pred. No.: 2.19e-129 Length: 491  
Score: 2413.50 Matches: 456  
Percent Similarity: 95.02% Conservative: 21  
Best Local Similarity: 90.84% Mismatches: 14  
Query Match: 84.98% Indels: 11  
DB: 7 Gaps: 3  
US-09-475-704A-4 (1-1509) x ADC72913 (1-491)  
QY 1 ATGGGCGCCGCGCCAGCATCTGCGCGCGAGAACTGAGCAAGTGAGAGAGATCCGC 60  
|||||

Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCCCGCGCGCAAGACACTACATGCTGAAAGCACTGTGTGGCCAGCCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisIleTrpMetIleLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGGAGCTTGGCCCTGAAACCCCGGCTGCTGAGACCGCCGAGGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGACAGCCCGCCCTGACAGCCGCGCAGAGAGAGCTGCGACCTGTACAA 240  
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuHisAsn 80  
QY 241 ACCGTGGCCACCCCTGTACTGCGTGCACCGCCGATCGAGTCCCGCGCAACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTrpCysValHisAlaGlyIleGluIleArgAspThrLysGluAla 100  
QY 301 CTGCAACAAGATCGAGAGAGAGAGAGACAAGTCCCAAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGluGluGluGluGluLysSerGlnGlnLysThrGlnGlnAlaLysGlu 120  
QY 361 GCCGAGCGCAAGGTGAGCCAGAACTACCCCATGCTGAGAACCTGCGAGGCGCCAGATGCTG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyIleMetVal 140  
QY 421 CACCAAGCCATCAGCCCGCCGACCCCTGAAAGCCTGGGTGAAGTGAATCGAGAGAGAGCC 480  
Db 141 HisGlnAlaLeuSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCCACCCCGCAGAG 540  
Db 161 PheSerProGluIleIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACAGATGTTGAACACCGTGGGCGGCCACAGCGCCGCGCATGAGTCTGAAGAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGCGCCGCGAGTGGAGCCGCTGCAACCCCGTGCAAGCCGCGCCGTG 660  
Db 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProAlaGlnAlaGlyProIle 220  
QY 661 GCCCGCGCGCAGATGCGCGCAGCCCGCGCGCAGCGCATCGCCGCGCCACAGCACCTG 720  
Db 221 AlaProGlyGlnMetLysArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGAGCAGATCGCCTGATGATGACCAACCAACCCCGCTGCCGTGGGCGCATCTACAAG 780  
Db 241 GlnGluGlnIleAlaTrpMetThrGlyAsnProProValProValGlyGluIleTrpLys 260  
QY 781 CGGTGATCATCTCGGCGCTGAACAAGATCGTGGCGATGTACAGCCCGCTGAGCATCTCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTrpSerProValSerIleLeu 280  
QY 841 GACATCCGCGCAGGCGCCCAAGAGCCCTTCGCGCACTACGTGAGCCGCTTCTTCAAGAC 900  
Db 281 AspIleLysGlnGlyProLysGluProPheArgAspTrpValAspArgPhePheLysThr 300  
QY 901 CTGCGCGCGCAGAGAGCCCAACCAAGCAAGTGAAGAACTGGATGACCAAGCCCTGCTGTG 960  
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAAAGCCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCGCCGCGCCACCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 GAGAGATGATGACCGCCTGCGCAGGCGGTGGCGGCGCCGCGCAAGAGCCCGCTGCTG 1080  
Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGAGCCAGCCCAAGCAAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
Db 361 AlaGluAlaMetSerGlnThrAsnAsn--SerIleLeuMetGlnArgSerAsnPheLys 379

```
QY 1141 GGCCCCCGCGCAAGTGAAGTGTCTCACTGGCGCAAGAGGGCCACATCGCCCAAGAAC 1200
    |||
    ::|||
Db 380 GlyPheLysArgThrValLysCysPheLysCysGlyGlyGluGlyHisIleAlaArgAsn 399
QY 1201 TGCCGCGCCCCCGCAAGAGAGGCTGTGGAAGTGGCGCAAGAGAGGGCCACCATGTAAG 1260
    |||
    |||
Db 400 CysArgAlaProArgLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 419
QY 1261 GACTGCAACCGAGCGCGCAAGCTTCTGGGCAAGATCTGGCCCAAGCACAAGAGGGCCG 1320
    |||
    |||
Db 420 AspCysThrGluArgGlnAlaAsnPheLysGlyLysIleTrpProSerHisLysGlyArg 439
QY 1321 CCGCGCAACTTCTGTGAGAACCGCAGCGAGCCCGCCGCCACCGTGGCCACCGCCCGC 1380
    |||
    |||
Db 440 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 453
QY 1381 CCGCGCAGAGCTTCCGCTTCGAGAGACCAACCCCGCCCAAGAGAGAGCCCAAGAC 1440
    |||
    |||
Db 454 ProAlaGlnSerPheArgPheGluGluThrThrProAlaLeuLysGlnGluGlnLysAsp 473
QY 1441 CGCGAGCCCTTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTGGGAGCGGCCCTG 1500
    |||
    |||
Db 474 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 489
QY 1501 AGCCAG 1506
    |||
    |||
Db 490 SerGln 491
```

## RESULT 12

ADN36396  
ID ADN36396 standard; protein; 514 AA.

AC ADN36396;

DT 15-JUL-2004 (first entry)

DE Human protein for anti-HIV vaccine.

KW anti-HIV; vaccine; HIV; promoter; viral particle; immunization.

OS Homo sapiens.

PN WO2004035006-A2.

PD 29-APR-2004.

PF 17-OCT-2003; 2003WO-US033112.

PR 18-OCT-2002; 2002US-0419465P.

PA (AARO-) AARON DIAMOND AIDS RES CENT.

PI Huang Y, Ho DD, Chen Z;

DR WPI; 2004-348328/32.

DR N-PSDB; ADN36395.

PT Nucleic acid vector comprising at least one HIV sequence operably linked to a promoter and encoding a protein that does not assemble into viral particles, useful in immunizing a subject against HIV infection.

PS Disclosure; SEQ ID NO 10; 166pp; English.

CC The invention relates to a nucleic acid vector comprising at least one HIV sequence operably linked to a promoter and encoding a protein that does not assemble into viral particles. The nucleic acid vector is useful in immunizing a subject against HIV infection. This sequence corresponds to a peptide used in the invention.

CC Sequence 514 AA;

Alignment Scores:

```
Pred. No.: 3,26e-129 Length: 514
Score: 2410.50 Matches: 457
Percent Similarity: 94.82% Conservative: 19
Best Local Similarity: 91.04% Mismatches: 15
Query Match: 84.88% Indels: 11
DB: 8 Gaps: 3
US-09-475-704A-4 (1-1509) x ADN36396 (1-514)
QY 1 ATGGGCGCCCGCGCCAGCATCTCTGCGCGGCGGAGAAAGCTGGACAAGTGGAGAAAGATCCGC 60
    |||
    |||
Db 24 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 43
QY 61 CTGGCGCCCGCGCGCAAGAGCACTACATGCTGAAGCAACCTGGTGTGGGCGGCGGAG 120
    |||
    |||
Db 44 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 63
QY 121 CTGGAGGGCTTCCGCTTGAACCCCGCGCTGCTGGAAGACCGCGGAGGCTGCAAGCATC 180
    |||
    |||
Db 64 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 83
QY 181 ATGAAGCAGCTGCGAGCCCGCGCTGCAAGCCGCGCATCGAGGTCCGACACCAAGAGGCC 240
    |||
    |||
Db 84 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuPheAsn 103
QY 241 ACCGTGGCCACCTCTGACTGCTGCGAGCGCGCATCGAGGTCCGACACCAAGAGGCC 300
    |||
    |||
Db 104 ThrValAlaThrLeuTyrCysValHisGluGlyIleGluIleArgAspThrLysGluAla 123
QY 301 CTGGACAAGATTCGAGAGAGAGAGAGACAAGTCCAGCAGAGAACCTGCAAGGCCCAAGAG 360
    |||
    |||
Db 124 LeuAspLysIleGluGluGlnAsnLysIleGlnGlnLysThrGlnGlnAlaLysLys 143
QY 361 GCCGACGCGCAAGGTGAGCCCACTACCCCATCGTGACAACCTGCAAGGCCCAAGAGGCC 420
    |||
    |||
Db 144 AlaAspGluLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMetVal 163
QY 421 CACCAGGCCATCAGCCCGCGCACCCCTGAGACGCTGGGTGAAGGTGATCGAGAGAGGCC 480
    |||
    |||
Db 164 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 183
QY 481 TTCAGCCCCCGAGGTATCCCATGTTCAACCGCCCTGAGCGGAGGCGGCCACCCCGAGAG 540
    |||
    |||
Db 184 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 203
QY 541 CTGAACACGATGTTGAACACCGGTGGCGCGCCACCAAGGCCCATGACAGATGCTGAAGGAC 600
    |||
    |||
Db 204 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 223
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGACCGCTGCAACCCCGCTGCAAGCGCGCGCGCTG 660
    |||
    |||
Db 224 ThrIleAsnGluGluAlaIleAlaGluTrpAspArgValHisProValHisIleGlyProIle 243
QY 661 GCGCGCGCGGAGATGCGCGGACCCCGCGCGAGCGAGCATGCGCGCGCCACCAAGACCCCTG 720
    |||
    |||
Db 244 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 263
QY 721 CAGAGCAGATCGCCTGATGACCAAGCAACCCCGCTGCGGCGCGGCGGCGCATCTACAAG 780
    |||
    |||
Db 264 GlnGlyGlnIleAlaTrpMetThrSerAsnProProValProValGlyGluIleTyrLys 283
QY 781 CGGTGATCATCTCGGCTGGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATCTCTG 840
    |||
    |||
Db 284 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 303
QY 841 GACATCCGCGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAGACCGGCTTCTCAAGAC 900
    |||
    |||
Db 304 AspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 323
QY 901 CTGCGCGCGGAGCAGGCGCCACCAAGAGAGCTGAAGACTGATGACCGGAGACCTGCTGCTG 960
    |||
    |||
Db 324 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 343
```

QY 961 CAGAACGCCAACCCCGACTGTCAGAACCATCTCGCGCTCTCGGCCCGCCACCCCTG 1020  
Db 344 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerIleu 363  
QY 1021 GAGGAGATGATGACCCGCTGCGCAGGGCGGTGGGGCGCCCGCCACAAGGCCCGGTGCTG 1080  
Db 364 GluGluMetThrAlaCysGlnGlyAlaGlyGlyProSerHisLysAlaArgValLeu 383  
QY 1081 GCCGAGCGATGAGCCAGGCCACAAGCGCTGAACATCATGATGACAGAGACCAACTTCAAG 1140  
Db 384 AlaGluAlaMetSerGlnAlaAsnGly---ThrIleLeuMetGlnArgSerAsnPheLys 402  
QY 1141 GGCCCCCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATGCGCAAGAAC 1200  
Db 403 GlySerLysArgIleValLysCysPheAsnCysGlyLysGlyGlyHisIleAlaArgAsn 422  
QY 1201 TGCCGCGCCCCCGCGCAAGAGGGTGTGCTGAAGTGCAGGAGAGGGCCACAGATGAAG 1260  
Db 423 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlyGlyHisGlnMetLys 442  
QY 1261 GACTGCACCGAGCGCGCAACTCTTGCGGCAAGATCTGGCCCAAGAGGGCGC 1320  
Db 443 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 462  
QY 1321 CCCGCAACTTCTGTGAGAACCGCAGCGAGCCCGCCGCCACCGTGCACCGCCCGC 1380  
Db 463 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 476  
QY 1381 CCCGCGAGAGCTTCCGCTTGCAGAGAACACCCCGCCCGCAAGAGAGGCCCAAGAC 1440  
Db 477 ProAlaGlnSerPheArgPheGluGluThrThrProAlaProLysGlnGluProLysAsp 496  
QY 1441 CGCAGACCTTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTCGAGAGCGGCCCTG 1500  
Db 497 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 512  
QY 1501 AGCCAG 1506  
Db 513 SerGln 514  
RESULT 13  
AAB69276  
ID AAB69276 standard; protein; 498 AA.  
XX AC AAB69276;  
XX 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX DE HIV-1 non-subtype B clone 96ZM751-3 gag protein.  
XX KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
XX KM vif; vpr; tat; rev; nef; vaccine.  
XX OS Human immunodeficiency virus 1.  
XX PN WO200026416-A1.  
XX PD 11-MAY-2000.  
XX PF 25-OCT-1999; 99WO-US024837.  
XX PR 02-NOV-1998; 98US-00184418.  
XX PA (UABR-) UAB RES FOUND.  
XX PI Hahn BH, Shaw GM, Gao F;  
XX DR WPI; 2000-365651/31.  
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.

XX Claim 41; Fig 14; 131pp; English.  
PS  
XX The present in invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 498 AA;  
Alignment Scores:  
Pred. No.: 7.64e-129 Length: 498  
Score: 2404.00 Matches: 448  
Percent Similarity: 94.22% Conservative: 25  
Best Local Similarity: 89.24% Mismatches: 25  
Query Match: 84.65% Indels: 4  
DB: 3 Gaps: 1  
US-09-475-704A-4 (1-1509) x AAB69276 (1-498)  
QY 1 ATGGCGCGCCGCGCAGCATCTCGCGCGCGAGAGCTGGAACAGTGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspGluTrpGluArgIleArg 20  
QY 61 CTGGCCCCGCGCGCAAGAGCACTACATGCTGAGACACCTGTGTGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisArgYrMetMetLysHisLeuIleTrpAlaSerArgGlu 40  
QY 121 CTGGAGGGCTTGCCCTGAACCCCGGCTGTGTGAGACCGCGAGGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATGAGCAGCTGACGCCCGCTGACAGCCGCGCAGAGAGCTGCGCAGCTGTACAAC 240  
Db 61 IleGlnGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuYrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGCCGCGCATCGAGGTCCGCGACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGluLysIleLysValArgAspThrLysGluAla 100  
QY 301 CTGGACAAGATCGAGAGAGAGAGACAAGTCCACGACAGAAGACCAGCAGCGCAGAG 360  
Db 101 LeuAspLysIleGluGluGlnAsnLysSerGlnLysIleGlnLysThrGluAla 120  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAGAACCTGACAGGGCCAGATG 420  
Db 121 ThrGlyGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACCAAGGCATCAGCCCCCGCACCTTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaIleTrpValLysValIleGluGluLysGly 160  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGGCCACCCCGCAGAC 540  
Db 161 PheAsnProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGGTGGGGCGCCACAGCGCCCATGCAATGCTGAAGGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGACCGCTGCACCCCGTGCAGAGCGCGCCCGTG 660  
Db 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCGCCCGCGCAGATGCGCGACCGCCCGCGGAGGACATCGCCGCGCCACAGACCCCTG 720  
Db 221 AlaProGlyGlnIleArgGluProArgGlySerAspIleAlaGlyThrThrGlyThrLeu 240  
QY 721 CAGAGACAGATCGCCTGATGACCAAGCAACCCCGGTGCCCGTGGGGCGACATCTACAAG 780

Db 241 GlnGlnGlnIleAlaIleTyrMetThrAsnAsnProProlleProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCTGAACAAGATCGCGATGTACAGCCCGTGAGCATCTG 840  
Db 261 ArgTrrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCCAGGGCCCCAAGAGAGCCCTTCCGCGACTACGTGAGACCCGCTTCTCAAGACC 900  
Db 281 AspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 300  
QY 901 CTGCGCGCCGAGCAGGCCACCCAGAGACTGTGAAGACTGGATGACCCGAGACCTGTGTG 960  
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLys\*\*\*TrpMetThrAspThrLeuVal 320  
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGCCCGCGCCACCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340  
QY 1021 GAGAGATGATGACCCGCTGCGAGGGGTGGGGCGGCCGACAAAGCCCGCTGCTG 1080  
Db 341 GluGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGAGCCAGGCCCAACAGCGTGAACATCATGATGACAGAAAGCACTTCAAG 1140  
Db 361 AlaGlnAlaMetSerGlnValAlaAsnThrAsnIleMetMetGlnLysSerAsnPhelLys 380  
QY 1141 GGCCCCCGCGCAAGCTCAAGTGTCTCACTGCGGCAAGAGGGCCACATCGCCAAAGAAC 1200  
Db 381 GlyProLysArgIleValLysCysPheAsnCysGlyArgGlnGlyHisIleAlaArgAsn 400  
QY 1201 TGCGCGCGCCCCCGCAAGAGGGCTGCTGAAGTGGCGCAAGAGGGCCACCATGAAG 1260  
Db 401 CysArgAlaProGlyLysLysGlyCysTrrLysCysGlyLysGlnGlyHisGlnMetLys 420  
QY 1261 GACTGCACCGGAGCGGCCCACTTCTTGGGCAAGATCTGGGCCAGCCACAAAGGCGCG 1320  
Db 421 AspCysThrGlnArgGlnAlaAsnPhelLeuGlyLysIleTrrProSerGlnLysGlyArg 440  
QY 1321 CCGGCGCACTTCTTCAGAACCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
Db 441 ProGlyAsnPhelLeuGlnAsnArgProGluProThrAlaProProAlaProThrAlaPro 460  
QY 1381 CCGCGCGAGAGCTTCCGCTTCGAGAGAGACCAACCCCGCGCGCGCGCGCGCGCGCG 1440  
Db 461 ProAlaGluSerPheArgPheGlnGluThrThrProAlaProArgGlnGlnLysAsp 480  
QY 1441 CGCGAGCCCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1500  
Db 481 -----LysGluProLeuThrAlaLeuLysSerLeuPheGlySerAspProLeu 496  
QY 1501 AGCCAG 1506  
Db 497 SerGln 498

RESULT 14  
AAB69275  
ID AAB69275 standard; protein; 494 AA.  
XX

AC AAB69275;  
XX  
DT 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX  
DE HIV-1 non-subtype B clone 962M651-8 gag protein.  
XX  
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
XX vif; vpr; tat; rev; nef; vaccine.  
OS Human immunodeficiency virus 1.  
XX  
PN WO200026416-A1.  
XX

PD 11-MAY-2000.  
XX  
PF 25-OCT-1999; 99WO-US024837.  
XX  
PR 02-NOV-1998; 98US-00184418.  
XX  
PA (UABR-) UAB RES FOUND.  
XX  
PI Hahn BH, Shaw GM, Gao F;  
XX  
DR WPI; 2000-365651/31.  
XX  
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.  
XX  
PS Claim 41; Fig 14; 131pp; English.  
XX  
CC The present invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 494 AA;  
XX  
Alignment Scores:  
Pred. No.: 9.92e-129 Length: 494  
Score: 2402.00 Matches: 455  
Percent Similarity: 93.85% Conservative: 18  
Best Local Similarity: 90.28% Mismatches: 19  
Query Match: 84.58% Indels: 12  
DB: 3 Gaps: 3  
US-09-475-704A-4 (1-1509) x AAB69275 (1-494)  
QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGGCAAGAGCTGGAAGTGGAGAAATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrrGlnLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGCAAGACACTACATCTGAAGCACCTGTGTGGCGCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysArgTrrMetIleLysHisLeuValTrrPalaserArgL 40  
QY 121 CTGAGGGGCTTCCGCTGAACCCCGGCGCTGTGAGAGCCCGGAGGGGCTGCAAGCAGATC 180  
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGCAGCGCGCGCTGCAGACCGGCGAGGAGCTGCGCAGCTGTACAAAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCTGTCAGCGCGCGCATCGAGGTCCGGACCAAGAGAGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnGlyValGlnValArgAspThrLysGlnAla 100  
QY 301 CTGAGACAGATCGAGGAGAGCAGAACTACCCCATCTGTGCAGAACTGCAGGGCGCAG 360  
Db 101 LeuAspArgIleGlnGlnGlnAsnLysIleGlnGlnLysIleGlnGlnLysThrGln 120  
QY 361 -----GCCGACGGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAGAACTGCAGGGCGCAG 414  
Db 121 GlnAlaAlaAspGlyLysValSerGlnAsnTrrProIleValGlnAsnLeuGlnGln 140  
QY 415 ATGTGTGACCAAGGCCATGACCCCGCGCAGCTGAACGCGCTGGGTGAAGTGTATGAGAG 474  
Db 141 MetValHisGlnLysLeuSerProArgThrLeuAsnAlaTrrValLysValIleGlnGln 160  
QY 475 AAGGCTTCAGCCCGGAGGTGATCCCATGTTCACCGCGCTGAGCGAGGGCGCCACCCCC 534

```
Db      161 LysAlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGluAlaThrPro 180
QY      535 CAGGACCTGAAACAGATGTTGAACACCGTGGGGCCGACAGGCCCGCATGAGATGCTG 594
Db      181 GlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeu 200
QY      595 AAGACACCATCAACGAGGAGGCCCGGAGTGGGACCGCTGCACCCCGTGACGCGGC 654
Db      201 LysAspThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGly 220
QY      655 CCGGTGGCCCCCGGACGATGCGCGACCCCGCGGACGAGACATGCGCGCGCCACCAAGC 714
Db      221 ProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSer 240
QY      715 ACCCTGAGAGAGACAGATGCTGTGGCTGAACAGAACCCCGCTGCGCGTGGGCGACATC 774
Db      241 ThrLeuGlnGluGlnIleAlaThrMetThrSerAsnProProIleProValGlyAspIle 260
QY      775 TACAACGGTGGATCATCTCTGGGCTTGAACAAGATCGTCGGATGTACAGCCCGTGAGC 834
Db      261 TyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSer 280
QY      835 ATCTGTGACATCCGCGGAGGCCCCAAGAGAGCCCTTCCGCACTACGTGAGACCGCTTCTC 894
Db      281 IleLeuAspIleLeysGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300
QY      895 AAGACCTGCGCGCGGACGAGCCACCGACGAGACGTGAAGATCGATGACCGAGACCTG 954
Db      301 LysThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeu 320
QY      955 CTGTGAGAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCTCTGCGCCCGCGCGCC 1014
Db      321 LeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGlyAla 340
QY      1015 ACCCTGAGAGAGATGATGACCGCTGCGAGGCGCGTGGGCGGCCCGGACAAAGCGCCG 1074
Db      341 ThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArg 360
QY      1075 GTGTGCGCGGAGCGGATGAGCCAGGCCAACAGCGTGAACATCATGATGACGAGAGAAC 1134
Db      361 ValLeuAlaGluAlaMetSerGlnThrAsnSerValAsnIleLeuMetGlnLysSerAsn 380
QY      1135 TTCAAGGGCCCCCGGCGCAAGTCAAGTCTTCACTGCGGCAAGAGAGGCCACATCGCC 1194
Db      381 PheLysGlyAsnLysAsnMetValLysCysPheAsnGlyLysGluGlyHisIleAla 400
QY      1195 AAGAACTGCGCGCGCCCGGCAAGAGGGCTGCTGAAGTGGCGGCAAGAGGCCACCAAG 1254
Db      401 ArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 420
QY      1255 ATGAAGACTGCAACCGAGCGCCAGGCCCACTTCTGGGCAAGATCTGGGCCACGACCAAG 1314
Db      421 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 440
QY      1315 GGCGCGCGCGGCAACTTCTGCAAGAACCGGACGCGCGCGCCCGGCAACCGTGCCCAAC 1374
Db      441 GlyArgProGlyAsnPheLeuGlnAsnArgProGlu-----ProThr 454
QY      1375 GCGCGCGCGCGCGGAGAGCTTCCGCTTCAAGAGAGACCAACCGCGCGCCCAAGAGAGCCC 1434
Db      455 AlaProProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGlnSer 474
QY      1435 AAGACCGCGGAGCGCTTACCGCGAGCGCCCTGACCGCCCTGCGGACGCTGTTCGAGCGGC 1494
Db      475 LysAsp-----ArgGluAlaLeuThrSerLeuLysSerLeuPheGlySerAsp 490
QY      1495 CCGCTGAGCCAG 1506
Db      491 ProLeuSerGln 494
```

RESULT 15  
ADC72875  
ID ADC72875 standard; protein; 508 AA.

```
XX      AC      ADC72875;
XX      DT      18-DEC-2003 (first entry)
XX      DE      HIV-1 gag protein containing an NCP sequence, SEQ ID NO 105.
XX      KW      complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;
XX      KW      human immunodeficiency virus; HIV.
XX      OS      Human immunodeficiency virus 1.
XX      PN      WO2003060098-A2.
XX      PD      24-JUL-2003.
XX      PF      10-JAN-2003; 2003WO-US000801.
XX      PR      11-JAN-2002; 2002US-0347369P.
XX      PA      (ACHI-) ACHILLION PHARM INC.
XX      PI      Beauchter D, Hou X, Marlor CW, Rice WG, Yang W;
XX      DR      WPI; 2003-646042/61.
XX      PT      Method of determining whether a compound inhibits formation of complex
XX      PT      between HIV nucleocapsid protein 7 polypeptide and HIV psi-site
XX      PT      oligonucleotide by comparing amount of complex formed in presence/absence
XX      PS      of compound.
XX      PS      Claim 2; SEQ ID NO 105; 105bp; English.
XX      CC      The invention relates to a novel method of determining whether a compound
XX      CC      inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)
XX      CC      polypeptide and an HIV-psi-site oligonucleotide. The method involves
XX      CC      adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and
XX      CC      the novel compound and comparing amount of complex formed, with that
XX      CC      formed in the absence of the compound, where a decrease in the amount of
XX      CC      complex formed in presence of the compound indicates that the compound
XX      CC      inhibits complex formation. The method of the invention involves anti-HIV
XX      CC      activity. The compounds identified are useful for treating a subject
XX      CC      infected with human immunodeficiency virus (HIV) by administering the
XX      CC      compound to the subject. This sequence represents an HIV-1 gag protein
XX      CC      which contains an NCP sequence of the invention.
XX      SQ      Sequence 508 AA;

Alignment Scores:
Pred. No.:      7,11e-128      Length:      508
Score:          2387.00      Matches:      454
Percent Similarity: 92.19%      Conservative: 18
Best Local Similarity: 88.67%      Mismatches: 26
Query Match:    84.05%      Indels:      14
DB:             7      Gaps:      3

US-09-475-704A-4 (1-1509) x ADC72875 (1-508)

QY      1      ATGGGCGCGCGCGGCGGACATCTCTGCGCGCGGAGAGCTGGAACAGTGGGAGAGATCCGC 60
Db      1      MetGlyAlaArgAlaSerIleLeuArgGlyThrLysLeuAspAlaTrpGluLysIleArg 20
QY      61      CTGCGCGCGCGGCGGCAAGACACTTCACTGCTGAAGACCTGTGTGGGCCAGCGCGGAG 120
Db      21      LeuArgProGlyGlyLysLysHisIleTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40
QY      121      CTGAGAGGCTTCCGCTTGAACCCCGGCTGCTGAGAGACCGCGGAGGCTGCAAGCAGATC 180
Db      41      LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysLysGlnIle 60
QY      181      ATGAAGCAGCTGCAAGCGCGCCCTGCAAGACCGGACGAGAGAGCTGCGGACGCTGTACAAC 240
Db      61      MetLysGlnLeuHisProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80
```

QY 241 ACCGTGGCCACCCTGTA CTGCTGCA CGCCGCATCGAGTCCGCGACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGluAsnIleLysValArgAspThrLysGluAla 100  
QY 301 CTGGACAAGATCGAGGAGGAGCAGAAC-----AAGTCCCGACAGAGAAGACCAGAGGCC 354  
Db 101 LeuAspLysIleGluGluGluGluAsnLysIleLysSerGlnGlnLysThrGlnGlnAla 120  
QY 355 AAGGAGCCGACGCGCAAGGTGAGCCGAACCTACCCCATCGTGCAAGACTGCAAGGCCAG 414  
Db 121 LysAlaAlaAspGluLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGln 140  
QY 415 ATGCTGCAACGAGCCATCAGCCCGCAGCCCTGAACGCTGGGTGAAGTGATCGAGAG 474  
Db 141 MetValHisGlnAsnLeuSerProArgThrLeuAsnAlaTyrValLysValIleGluGlu 160  
QY 475 AAGGCTTCAGCCCGCAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCACCCCC 534  
Db 161 LysAlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGluAlaThrPro 180  
QY 535 CAGAACCTGAACACGATGTTGAACACCGTGGCGCGCCACAGCGCCCATGCAAGTCTG 594  
Db 181 GlnAspLeuSerThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeu 200  
QY 595 AAGGACACCATCAACAGAGAGGCGCGCAGTGGACCGCTGCACCCCGTGACAGCGCGC 654  
Db 201 LysAspThrIleAsnGluGluAlaIleGluTyrAspArgLeuHisProValHisAlaGly 220  
QY 655 CCGGTGGCCCCCGCGCAGATGCGCGACCCCGCGCGCAGCATCGCGCGCGCACCCAGC 714  
Db 221 ProMetAlaProGlyGlnLeuArgLupProArgLysSerAspIleAlaGlyThrThrSer 240  
QY 715 ACCCTGCAGAGACAGATCGCCTGATGACCAAGCAACCCCGCTGCGCCGTGGCGCATC 774  
Db 241 ThrLeuArgGluGlnIleAlaTyrMetThrSerAsnProProIleProValGlyAspIle 260  
QY 775 TACAAGCGGTGATCATCTGCGCTGAACAACAAGATGCGGATGTACAGCCCGGTGAGC 834  
Db 261 TyrLysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSer 280  
QY 835 ATCTTGACATCCGCGCAGGCGCCCAAGAGAGCCCTCCGCGACTACGTGAGCCGCTTCTC 894  
Db 281 IleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePhe 300  
QY 895 AAGACCTCGCGCGCGCAGCAGCGCCACCCAGGACCGTGAAGAAGACTGATGACCGAGACCCTG 954  
Db 301 LysAlaLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTyrMetThrGluThrLeu 320  
QY 955 CTGGTGCAAGACGCGCAACCCCGACTGCAAGACCATCTCGCGCTCTCGCGCCCGCGCGCC 1014  
Db 321 LeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyIleGlyAla 340  
QY 1015 ACCCTGGAGGAGATGATGACCGCTGCGCAGGCGCGTGGCGCGCCCGCGCACAGGCCCGC 1074  
Db 341 ThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArg 360  
QY 1075 GTGCTGGCGCGAGCGCATGAGCCAGGCGCAACAGCGGTGAACATCATGATGCAAGAAGCAAC 1134  
Db 361 ValLeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetMetGlnArgSerAsn 380  
QY 1135 TTCAAGGGCCCCCGCGCAACGTTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCC 1194  
Db 381 PheLysSerSerLysArgIleValLysCysSerAsnCysGlyLysGluGlyHisIleAla 400  
QY 1195 AAGAACTGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGCGGCAAGAGGGCCACAG 1254  
Db 401 ArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 420  
QY 1255 ATGAAGGACTGCACCGCAGCGCCAGGCGCAACTTCTGCGGCAAGATGCGCCAGCCACAAG 1314  
Db 421 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 440

QY 1315 GGCGCGCCCGGCAACTTCTGTGCAGAACCGCAGCGAGCCCGCGCGCCCGCACCGTG----- 1368  
Db 441 GlyArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSer 460  
QY 1369 -----CCACCGCGCGCGCGCGCGAGAGCTTCCGCTTCGAGGAGACC 1410  
Db 461 PheArgAsnArgProGluProThrAlaProProAlaGluSerPheArgPheGluGluThr 480  
QY 1411 ACCCGCGCGCGCGCAAGCAGAGCGCCGAAGACCGCGAGCCCTTACCGCGAGCGCCCTGAACCGCC 1470  
Db 481 ThrProThrProLysGlnGluProLysAsp-----ArgAspProLeuThrSer 496  
QY 1471 CTGGCGAGCTGTTCGGCAGCGCGCGCCCTGAGCCAG 1506  
Db 497 LeuLysSerLeuPheGlySerAspProSerSerGln 508

Search completed: March 11, 2005, 15:22:46  
Job time : 241.806 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 14:56:45 ; Search time 259.58 Seconds  
(without alignments)  
5953.668 Million cell updates/sec

Title: US-09-475-704A-4  
Perfect score: 2840  
Sequence: 1 atggcgccgcgcgcgcagcat.....gcgcgcgcctgagccagtaa 1509

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues  
Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US09475704/runat\_10032005\_140221\_14821/app\_query.fasta\_1.3342  
-DB=uniprot\_03 -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704\_QCGN\_1\_1\_614\_@runat\_10032005\_140221\_14821 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : uniprot\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2661	93.7	502	2	Q9WF78	Q9wf78 human immun
2	2464	86.8	492	2	Q901X9	Q901x9 human immun
3	2458	86.5	500	2	Q90056	Q90056 human immun
4	2457	86.5	492	2	Q8AFW4	Q8afw4 human immun
5	2445.5	86.1	495	2	Q6XKD2	Q6xkd2 human immun
6	2440	85.9	492	2	Q8UTX7	Q8utx7 human immun
7	2439.5	85.9	499	2	Q8USV8	Q8usv8 human immun
8	2439	85.9	508	2	Q8USY5	Q8usy5 human immun
9	2438.5	85.9	497	2	Q6S7X9	Q6s7x9 human immun
10	2437	85.8	492	2	Q90ME1	Q90me1 human immun
11	2435.5	85.8	491	2	Q9QN98	Q9qn98 human immun
12	2433	85.7	492	2	Q901X0	Q901x0 human immun
13	2430.5	85.6	491	2	Q7SLX7	Q7slx7 human immun
14	2430	85.6	492	2	Q901Z7	Q901z7 human immun
15	2430	85.6	492	2	Q90ME2	Q90me2 human immun
16	2429	85.5	492	2	Q6S859	Q6s859 human immun

17	2429	85.5	500	2	Q8UT80	Q8ut80 human immun
18	2428.5	85.5	491	2	Q7SLY1	Q7slY1 human immun
19	2427.5	85.5	491	2	Q90090	Q90090 human immun
20	2427	85.5	492	2	Q8UTK6	Q8utK6 human immun
21	2425.5	85.4	491	2	Q8AFI1	Q8afI1 human immun
22	2425	85.4	488	2	Q7SLY2	Q7slY2 human immun
23	2424	85.4	500	2	Q90ME7	Q90me7 human immun
24	2422.5	85.3	491	2	Q90081	Q90081 human immun
25	2421.5	85.3	491	2	Q90064	Q90064 human immun
26	2421.5	85.3	491	2	Q7SLY0	Q7slY0 human immun
27	2421.5	85.3	491	2	Q7SLY6	Q7slY6 human immun
28	2421.5	85.3	491	2	Q9Q3C6	Q9q3C6 human immun
29	2419	85.2	492	2	Q90MF5	Q90mf5 human immun
30	2419	85.2	492	2	Q9WF93	Q9wf93 human immun
31	2417.5	85.1	501	2	Q8UTD4	Q8utD4 human immun
32	2417	85.1	492	2	Q9WF90	Q9wf90 human immun
33	2417	85.1	500	2	Q8UTN3	Q8utN3 human immun
34	2416.5	85.1	499	2	Q994H3	Q994h3 human immun
35	2415	85.0	492	2	Q6S7X0	Q6s7X0 human immun
36	2415	85.0	492	2	Q6S883	Q6s883 human immun
37	2414.5	85.0	495	2	Q8USZ4	Q8usz4 human immun
38	2414.5	85.0	499	2	Q6X6H4	Q6x6h4 human immun
39	2414	85.0	492	2	Q90MF4	Q90mf4 human immun
40	2414	85.0	492	2	Q9WFA1	Q9wfA1 human immun
41	2413.5	85.0	491	2	Q90072	Q90072 human immun
42	2413.5	85.0	499	2	Q8UTH9	Q8uth9 human immun
43	2413	85.0	492	2	Q9WF63	Q9wf63 human immun
44	2412.5	84.9	503	2	Q7SQ46	Q7sq46 human immun
45	2412	84.9	492	2	Q9WF87	Q9wf87 human immun

ALIGNMENTS

RESULT 1  
ID Q9WF78 PRELIMINARY; PRT; 502 AA.  
AC Q9WF78;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Gag polypeptide.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_Taxid=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vanberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RL J. Virol. 73:4427-4432(1999).  
DR EMBL; AF110967; AAD17071.1; -.  
DR HSSP; P24736; INCP.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.

DR PROSITE; PS50158; ZF\_CCHC; 2.  
KM AIDS: Core protein; Metal-binding; Polyprotein; Zinc; Zinc-finger.  
SQ SEQUENCE 502 AA; 55969 MW; E3ACPIF712A0C50F CRC64;

Alignment Scores:

Pred. No.:	6.42e-107	Length:	502
Score:	2661.00	Matches:	502
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	93.70%	Indels:	0
DB:	2	Gaps:	0

US-09-475-704A-4 (1-1509) x Q9W7F8 (1-502)

```
QY 1 ATGGGCGCCGCGCCAGCATCTGCGGCGGAGAAAGTGGACAAGTGGAGAGAATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20
QY 61 CTGCGCCCGCGCGGCAAGAAAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysIleTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40
QY 121 CTGAGGGGCTTGCCCTGAAACCCCGGCTGTGAGACCGCCGAGGCTGCAAGCAGATC 180
Db 41 LeuGluGlyPheAlaLeuAsnProGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60
QY 181 ATGAAGCAGCTGCAGCGCCGCGCTGAGACCGGCGAGGAGGCTGGCAGCCCTGTACAAC 240
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCCCTGTACTGCGTGCAGCGCGGCGATCGAGGTCCGCGACACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluValArgAspThrLysGluAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGAGAGAAAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 101 LeuAspLysIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCTGTCAGAACTGCGAGGCGCAGATGTG 420
Db 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140
QY 421 CACCAAGGCCATCAGCCCCCGGAGCCCTGAAAGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCGCAGAGAC 540
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY 541 CTGAACACGATGTTGAACACCGTGGGGCGGCGCAGGCGCCAGTGCAGATGTGAAGAC 600
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGACCGGCTGCACCCCGTGCAGAGCGCGCCGCTG 660
Db 201 ThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValGlnAlaGlyProVal 220
QY 661 GCCCGCCGCGCAGATGCGCGAGACCCCGCGGCGAGGACATGCGCGCGCCAGCAGACCCCTG 720
Db 221 AlaProGlyGlnMetArgAspProArgGlySerAspIleAlaGlyAlaThrSerThrLeu 240
QY 721 CAGGAGCAGATCGCCTGGATGACAGCAACACCCCGCTGCGCTGGGGGAGACATCTTAAG 780
Db 241 GlnGluGlnIleAlaIleTrpMetThrSerAsnProProValProValGlyAspIleTyrLys 260
QY 781 CGGTGATCATCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTCTG 840
Db 261 ArgTrpIleIleLeuGlyLysLeuAsnLysIleValArgMetLysSerProValSerIleLeu 280
QY 841 GACATCCGCGCAGGGGCCCAAGAGAGCCCTTCCGCGACTACGTGAGACCGCTTCTCAAGACC 900
Db 281 AspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePheLysThr 300
```

```
QY 901 CTGCGCGCCGAGAGAGGCCACCCAGAGAGCGTGAAGAACTGATGACCGAGACCTGTGTG 960
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrGluThrLeuLeuVal 320
QY 961 CAGAAAGCCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGCCCGCGCCAGCCCTG 1020
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340
QY 1021 GAGGAGATGATGACCGCCTGCGAGGGCGGTGGCGGCGCCCGGCGCAAGGCGCGTGTG 1080
Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360
QY 1081 GCCGAGCGGATGAGCCAGGCGCAACAGCGTGACATCATGATGACAGAGAGCACTTCAAG 1140
Db 361 AlaGluAlaMetSerGlnAlaAsnSerValAsnIleMetMetGlnLysSerAsnPhelys 380
QY 1141 GGCCTCCCGCGCAACGTCAAGTGTCTCAACTGCGGCGCAAGAGGGCCACATCGCCAAGAAC 1200
Db 381 GlyProArgArgAsnValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLysAsn 400
QY 1201 TGCCCGCGCCCGCGCAAGAGGGCTGTGAAAGTGCAGGAGAGGGCCACAGATGAAG 1260
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420
QY 1261 GACTGCACCGAGCGCCAGAGCCCACTTCTGCGCAAGATCTGGCCAGCCAGCCAGAGGGCGC 1320
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGCGCAACTTCTTGCAAGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 441 ProGlyAsnPheLeuGlnAsnArgSerGluProAlaAlaProThrValProThrAlaPro 460
QY 1381 CCGCGCGAGAGCTTCCGCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 461 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProLysAsp 480
QY 1441 CGCGAGCCCTACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 481 ArgGluProTyrArgGluProLeuThrAlaLeuArgSerLeuPheGlySerGlyProLeu 500
QY 1501 AGCCAG 1506
Db 501 SerGln 502
```

RESULT 2

Q901X9 PRELIMINARY; PRT; 492 AA.

AC Q901X9;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Gag protein.

GN Name=gag;

OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2156558; PubMed=11709097; DOI=10.1089/0892220152644232;

RA van Harmelen J., Williamson C., Kim B., Morris L., Carr J.,

RA Abdool Karim S.S., McCutchan F.;

RT "Characterization of full-length HIV type 1 subtype C sequences from

RT South Africa.";

RL AIDS Res. Hum. Retroviruses 17:1527-1531 (2001).

RN [2]

RP SEQUENCE FROM N.A.

RA van Harmelen J.H., Williamson C., Kim B., Morris L., Carr J.,

RA Abdool Karim S.S., McCutchan F.;

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY043175; AAL05325.1; -.

DR PIR; S49086; S49086.

DR HSSP; Q70622; 1HVN.

DR GO: 0019012; C:virion; IEA.  
DR GO: 0003676; F:nucleic acid binding; IEA.  
DR GO: 0005198; F:structural molecule activity; IEA.  
DR GO: 0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zF-CHC; 2.  
DR PRINTS; PR00939; C2HCNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; ZnF\_C2HC; 2.  
DR PROSITE; PS50158; zF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polyprotein; zinc; zinc-finger.  
SQ SEQUENCE 492 AA; 55196 MW; CFPAB0D5BA9FABE9 CRC64;

## Alignment Scores:

Pred. No.:	1.91e-98	Length:	492
Score:	2464.00	Matches:	466
Percent Similarity:	95.62%	Conservative:	14
Best Local Similarity:	92.83%	Mismatches:	12
Query Match:	86.76%	Indels:	10
DB:	2	Gaps:	2

US-09-475-704A-4 (1-1509) x Q901X9 (1-492)

QY 1 ATGGGCGCCGCGCGCGCATCTCTGGCGCGGAGAGCTGGACAAGTGGAGAGATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGlyLysIleArg 20  
QY 61 CTGGCGCCGCGCGCGCAAGAGCACTATGCTGAAGCACTGTGTGGCCAGCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysHisIleTyrMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAGGGCTTCGCGCTGAACCCCGCGCTGTGTGAAGCCGCGGAGGCTGCAAGAGATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGACCGCGCGCTGTGACCGCGAGAGAGAGTGGCAGCGCTGTACAAC 240  
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGAAGCGCGGATCGAGTCCGCGACACCAAGAGAGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGAGACAAGTCCCGACAGAGAGAGAGAGAGAGAGAG 360  
DB 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnAlaLysAla 120  
QY 361 GCGGACGCGAAGTGAAGCCAGAACTAACCCCATCTGTGACAGAACTGAGGCGCAATGTG 420  
DB 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACGAGGCCATAGCCCCCGGACACCTGAAGCGCTGGGTGAAGTGAAGTGAAGAGAGGCC 480  
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCAACCGCTGAGCGAGGCGGCCACCCCGCAGGAC 540  
DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACAAGATGTTGAACACCGTGGCGCGGCGACCGCGCGCATGAGTGTGAAGAGAC 600  
DB 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGCGCGCGGAGTGGGACCGCGCTGCACCCCGTGCAGGCGCGCGCGTG 660  
DB 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgValHisProValHisAlaGlyProIle 220  
QY 661 GCGCCCGCGCGAAGTGGCGGACCGCGCGCGGAGCATCGCGCGCGCGCACAGACCCCTG 720  
DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGAGCAGATCGCCTGTGATGACCAACAACCCCGCGTCCCGTGGCGGACATCTACAAG 780

DB 241 GlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATCTCTG 840  
DB 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGAGGCGCCCAAGAGAGCCCTTCGCGCACTACGTGACCGCTTCTTCAAGACC 900  
DB 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValaLysArgPhePheLysThr 300  
QY 901 CTGGCGCGCGAGAGCGCCACCGAGAGCTGAAGAACTGGATGACCGAGACCTGTGTGTG 960  
DB 301 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAAAGCCCAACCCCGACTGCAAGACCATCTCGCGCGCTCTCGGCGCGCGCGCACCTG 1020  
DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGGAGGCGGTGGCGCGCGCGCGCGCGCGCGCGCTGTG 1080  
DB 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360  
QY 1081 GCGGAGCGGATGAGCCAGGCGCAAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140  
DB 361 AlaGluAlaMetSerGlnAlaAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380  
QY 1141 GCGCCCGCGCGCAACGTCAAGTCTTCAACTGCGGCGCAAGAGGCGCACATCGCGCAAGAC 1200  
DB 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 400  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGGAGAGGCGCGCACAGATGAAG 1260  
DB 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420  
QY 1261 GACTGCACCGAGCGCGCGCAACTTCTGGGCAAGATCTGGCGCGCACAGAGGCGCGC 1320  
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
QY 1321 CCGGCGCACTTCTGCAAGAACCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
DB 441 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454  
QY 1381 CCGCGCGAGAGCTTCGCTTGAAGAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGAGAC 1440  
DB 455 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProIleGlu 474  
QY 1441 CGCGAGCCCTACCGCGAGCGCGCTGACCGCGCTGTGCGAGCGCGCGCGCGCGCTG 1500  
DB 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
QY 1501 AGCCAG 1506  
DB 491 SerGln 492  
RESULT 3  
ID 090056 PRELIMINARY; PRT; 500 AA.  
AC 090056;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9102570; PubMed=9847317;  
RA Iole K.S., Bollinger R.C., Paranjape R.S., Gaskart D., Kulkarni S.S.,  
RA Novak N.G., Ingersoll R., Sheppard H.W., Ray S.C.;  
RT "Full-length human immunodeficiency virus type 1 genomes from subtype

RT C-infected seroconverters in India, with evidence of intersubtype  
RT recombination."  
RL J. Virol. 73:152-160 (1999).  
DR EMBL; AF067154; AAD12071.1; -.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; F:viral life cycle; IEA.  
DR InterPro; IPR00721; Gag\_P24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; t2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.  
SQ SEQUENCE 500 AA; 55878 MW; 25AA23024004E325 CRC64;

Alignment Scores:

Pred. No.: 3.46e-98 Length: 500  
Score: 2458.00 Matches: 465  
Percent Similarity: 95.45% Conservative: 17  
Best Local Similarity: 92.08% Mismatches: 15  
Query Match: 86.55% Gaps: 8  
DB: 2 Gaps: 3  
US-09-475-704A-4 (1-1509) x O90056 (1-500)

QY 1 ATGGGCGCCCGCGCGCGCATCTCTGGCGGCGAGAGCTGGACAAGTGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyLysLeuAspLysTrpGluArgIleArg 20  
QY 61 CTGGCGCCCGCGCGCGAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgL 40  
QY 121 CTGGAGGGGCTTCGCGCTGAACCCCGGCTGTGTGAAGACCGCGGAGGGCTGAAGCATC 180  
Db 41 LeuAspArgPheAlaLeuAsnProGlyLeuLeuGluAlaAlaGluGlyCysLysGlnI 60  
QY 181 ATGAAGCAGCTGCAAGCCCGCTGCGAGACCGGCACTGCGAGCTGCGAGCTGTACAAC 240  
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLeuArgSerLeuHisAsn 80  
QY 241 ACCGTGGCCACCTCTGTACTGCGGTGCACGCGCGCATGAGGTCCGCGACCAAGAGAGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGluGlyIleGluValArgAspThrLysGlnAla 100  
QY 301 CTGCAACAAGATGAGAGAGAGAGAGAAACAAGTCCAGCAGAGAACCCAGCGCAAGAGAG 360  
Db 101 LeuAspLysIleGluGluGlnGlnAsnLysSerGlnGlnLysThrGlnAlaLysGlu 120  
QY 361 GCCGAGCGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACTGCAAGGCGCAGATGTG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACCAAGGCATCAGCCCGCGCACCTGAACGCGCTGGTGAAGGTGATCGAGAGAGAGCC 480  
Db 141 HisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLysAla 160  
QY 481 TTCAGCCCCGAGGTGATCCCCATGTTACCGCGCTGAGCGAGGGCGCCACCCCCAGAGAC 540  
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACAAGATGTTGAACACCGTGGCGGCGCACCGCGCATGCAAGTCTGAAGGAC 600

Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGAGACCGGCTGCACCCCGGTGCAGCGCCCGCGTG 660  
Db 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCCCGCGCGCAGATGCGGAGCCCGCGCGGAGGAGCATCGCGCGCGCCAGCACACCTG 720  
Db 221 AlaProGlyGlnIleArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGACGAGATGCGCTGTGATGACCAACACCCCGGTGGCGGTGGCGGACATCTACAAG 780  
Db 241 GlnGluGlnIleAlaTrpMetThrGlyAsnProProValProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCGCTGAACAAGATCGTGGAGTGTACAGCCCGTTCATCAGACC 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGAGGCGCCCAAGAGACCCCTCCGCGACTACGTGAGACCGCTTCTCAAGACC 900  
Db 281 AspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPheArgThr 300  
QY 901 CTGCGCGCGGAGCAGGCGCACCCAGAGACGTGAAGAACTGATGACCGAGACCTGTGCTG 960  
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAAGCCCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCCACCTGTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 GAGGAGATGATGACCGCTGTCCAGAGGCGGTGGGCGCGCCGCGCAAGGCGCGGTGCTG 1080  
Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360  
QY 1081 GCCGAGGCGATGAGCCAGCCCAACAGCCGTGAACATCATGATGATGAGAGAGCAACTTGA 1140  
Db 361 AlaGluAlaMetSerGlnAlaAsnSer--AsnIleLeuMetGlnArgSerAsnPheLys 379  
QY 1141 GGGCGCGCGCAACGTCAGAGTGTCTTCACTGCGGCGCAAGAGGCGCCACATCGCCAAGA 1200  
Db 380 GlySerLysArgThrValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 399  
QY 1201 TGCGCGCGCGCGCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGCGCCACAGATGAAG 1260  
Db 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 419  
QY 1261 GACTGCACCGAGCGCGCAAGGCTTCTTGGGCAAGATCTGGCCAGCCACCAAGGCGCGC 1320  
Db 420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCGGCGCACTTCTTGAGAAACCGCAGGAGCCCGCGCGCCACCGTG-----CCC 1371  
Db 440 ProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaArgProGluPro 459  
QY 1372 ACCGCGCGCGCGCGAGAGCTTCGCTTGCAGAGAGACCAACCCCGCGCCCAAGCAGAG 1431  
Db 460 ThrAlaProProAlaGluSerPheArgPheGluGluThrThrProAlaLeuLysGlnGlu 479  
QY 1432 CCAAGAGACCGGAGCCCTACCGCGAGCCCTGACCGCCCTGCGAGCCTGTTCGCGAGC 1491  
Db 480 ProlLysAsp-----ArgGluProLeuThrSerLeuLysSerLeuPheGlySer 495  
QY 1492 GGGCGCGCTGAGCCAG 1506  
Db 496 AspProLeuSerGln 500

RESULT 4  
Q8AFW4 PRELIMINARY; PRT; 492 AA.  
ID Q8AFW4;  
AC Q8AFW4;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)

DE	Gag protein.
GN	Name=gag;
OS	Human immunodeficiency virus 1.
OC	Viruses; Retroid viruses; Retroviridae; lentivirus.
OX	NCBI_TaxID=11676;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22525961; PubMed=12639249; DOI=10.1089/088922203762688649;
RA	Williamson C., Morris L., Maughan M.F., Ping L.-H., Dryga S.A.,
RA	Thomas R., Reap E.A., Chlifiers T., van Harmelen J., Pascual A.,
RA	Ramjee G., Gray G., Johnston R., Karim S.A., Swanson R.;
RT	"Characterization and selection of HIV-1 subtype C isolates for use
RT	vaccine development.";
RL	AIDS Res. Hum. Retroviruses 19:133-144(2003).
DR	EMBL; AF544010; AA016234.1; -.
DR	HSSP; Q70622; 1HVN.
DR	GO; GO:0019012; C:vitrion; IEA.
DR	GO; GO:0003676; F:nucleic acid binding; IEA.
DR	GO; GO:0005198; F:structural molecule activity; IEA.
DR	GO; GO:0016032; P:viral life cycle; IEA.
DR	InterPro; IPR000721; Gag_p24.
DR	InterPro; IPR010999; Retrovir_matrix.
DR	InterPro; IPR000071; Retrovir_p17.
DR	InterPro; IPR008916; Retrov_capsid_C.
DR	InterPro; IPR008919; Retrov_capsid_N.
DR	InterPro; IPR001878; Znf_CCHC.
DR	Pfam; PF00540; Gag_p17; 1.
DR	Pfam; PF00607; Gag_p24; 1.
DR	Pfam; PF00098; zf_CCHC; 2.
DR	SMART; SM00343; Znf_C2HC; 2.
DR	PROSITE; PS0158; ZF_CCHC; 2.
KW	AIDS; Core protein; Polyprotein.
SEQ	SEQUENCE 492 AA; 55221 MW; B1922020A1AE861B CRC64;

### Alignment Scores:

Pred. No.:	3.83e-98	Length:	492
Score:	2457.00	Matches:	465
Percent Similarity:	95.22%	Conservative:	13
Best Local Similarity:	92.63%	Mismatches:	14
Query Match:	86.51%	Indels:	10
DB:		Gaps:	2

US-09-475-704A-4 (1-1509) x Q8AFW4 (1-492)

Qy	1	ATGGCGCCCGCGCCGACATCTTCGCGGCGAGAGCTGACAAGTGGAGAAATCCGC	60
Db	1	MetGlyAlaArgAlaSerIleLeuArgGlyGluLeuAspLysTrpGluLysIleArg	20
Qy	61	CTGGCCCCCGCGCGGCAAGACACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG	120
Db	21	LeuArgProGlyGlyLysHisIleTyrMetLeuLysHisIleValTrpAlaSerArgGlu	40
Qy	121	CTGAGGGGCTTCGCCCTGAACCCCGGCTGCTGAGACCGCCGAGGCTGCAAGCAGATC	180
Db	41	LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle	60
Qy	181	ATGAGCAGCTGCAGCCCGCCCTGCAGACCGGCACCGAGAGCTGCGCAGCTGTACAC	240
Db	61	MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn	80
Qy	241	ACCGTGGCCACCCCTGTACTGCGTGACGCGCGCATCGAGTCCGCGACCAAGAGAGCC	300
Db	81	ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla	100
Qy	301	CTGCAACAAGATCGAGAGGAGCAGAAACAATCCAGCAGAAAGCCAGCGCCAAGAG	360
Db	101	LeuAspLysIleGlnGluGlnGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla	120
Qy	361	GCCGACGGCAAGGTAGACCCAGAACTACCCCATCGTGCAGAACTCTGCAGGGCCAGATG	420
Db	121	AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal	140
Qy	421	CACCAAGGCCATCAGCCCCCGCACCCCTGAACGCGCTGGGTGAAGGTGATCGAGGAGAGGCC	480

Db	141	HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValIleGluGluIleValAla	160
QY	481	TTCAAGCCCCGAGGTGATCCCATGTTACAGCCCTGAGCGAGGCGGCCACCCCGCAGAC	540
Db	161	PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp	180
QY	541	CTGAACACGATGTTGAACACCGTGGCGGCCACAGCGCCCATGCAGATGCTGAAGAC	600
Db	181	LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuIleAsp	200
QY	601	ACCATCAACGAGGAGCGCCCGAGTGGGACCGGCTGCACCCCGTGCAGCGCGGCCCGTG	660
Db	201	ThrIleAsnGlu**AlaIaGluTrpAspArgLeuHisProValHisIaGlyProIle	220
QY	661	GCCCCCGCGCAGATGCGCGCACCCCGCGCGCAGCAGCATCGCCGCGGCCACAGACCTG	720
Db	221	AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu	240
QY	721	CAGAGACAGATCGCCTGATGACCAAGACCCCGCGCGTGGCCGTGGCGCATCTTAAG	780
Db	241	GInGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyrIys	260
QY	781	CGGTGATGATCCTGGGCTGAAACAAGATCGTGGGATGTAACGCCCGTGAGCATCCTG	840
Db	261	ArgTrpIleIleLeuGlyLeuAsnIleValArgMetTyrSerProValSerIleLeu	280
QY	841	GACATCCGCGCAGGGCCCGCAAGAGCCCTTCGCGACTAGTGAGCCGCTTCTTAAGAC	900
Db	281	AspIleArgGlnGlyProIyGluProPheArgAspTyrValAspArgPhePheIyThr	300
QY	901	CTGCGCGCGCAGCAGGCCCAACCAAGACGTGAAGAACTGATGACCGAGACCTGCTGTG	960
Db	301	LeuArgAlaGluGlnAlaThrGlnGluValIyAsnTrpMetThrAspThrLeuLeuVal	320
QY	961	CAGAACGCCAACCCCGACTGCAAGACCATCCTGCGCGCTCTGCGCCCGCGCCACCTG	1020
Db	321	GlnAsnAlaAsnProAspCysIyThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu	340
QY	1021	GAGGATGATGACCGCTGCGCAAGCGCTGGCGCGGCCCGCCACAAGCGCGCTGTG	1080
Db	341	GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGluHisIyAlaArgValLeu	360
QY	1081	GCCGAGCGGATGAGCCAGGCCAACAGCGTGAACATCATGATGCAGAAAGCAACTTGAAG	1140
Db	361	AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPheIyAs	380
QY	1141	GCCCCCGCGGCAACGTCAAGTGTCTTAACGTGCGCAAGAGGCGCCACATGCGCAAGAC	1200
Db	381	GlyProArgArgIleValIyAsCysPheAsnIyGlyIyGluGlnHisIleAlaArgAsn	400
QY	1201	TGCCGCGCGCCCGCAAGAGGGCTGCTGAAGTGCAGCAAGAGGCGCCACGATGAG	1260
Db	401	CysArgAlaProArgIyLeuIyGlyCysTrpIyAsCysGlyIyGluGlnHisIyGlnMetIyAs	420
QY	1261	GACTGCACCGAGCGCCAGGCCAATTCTTGGGCAAGATCTGCGCCAGCAAGAGGCGCC	1320
Db	421	AspCysThrGluArgGlnAlaAsnPheLeuGlyIyIleTrpProSerHisIyAsGlyArg	440
QY	1321	CCCGGCAACTTCTGCAGAACCGCAGCGCGCGCCCGCCACCGTGCACCGCGCC	1380
Db	441	ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro	454
QY	1381	CCCGCGAGAGCTTCGCTTGCAGAGAGACACCCCGCCCAAGAGAGAGCCCAAGAC	1440
Db	455	ProAlaGluSerPheArgPheGluGlnThrThrProAlaProIyGlnGluProIleGlu	474
QY	1441	CGCGAGCCCTAACCGGAGCGCCCTGCAGCGCGCTGCGAGCCTGTTGGGAGCGCGCCCTG	1500
Db	475	-----ArgGluProLeuThrSerLeuIySerLeuPheGlySerAspProLeu	490
QY	1501	AGCCAG 1506	

Db 491 SerGln 492

RESULT 5  
Q6XKD2 PRELIMINARY; PRT; 495 AA.  
ID Q6XKD2  
AC Q6XKD2;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14686322; DOI=10.1089/088922203322588396;  
RA Papathanasopoulos M.A., Patience T., Meyers T.M., McCutchan F.E.,  
RA Morris L.;  
RT "Full-length genome characterization of HIV type 1 subtype C isolates  
RT from two slow-progressing perinatally infected siblings in South  
RT Africa."  
RL AIDS Res. Hum. Retroviruses 19:1033-1037(2003).  
DR EMBL; AY228556; AAO65554.1; -.  
DR HSSP; P12493; 1AK4.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf\_CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.  
SQ SEQUENCE 495 AA; 55437 MW; 480D89863166ABCB CRC64;

Alignment Scores:  
Pred. No.: 1.19e-97 Length: 495  
Score: 2445.50 Matches: 458  
Percent Similarity: 95.22% Conservative: 20  
Best Local Similarity: 91.24% Mismatches: 17  
Query Match: 86.11% Indels: 7  
DB: 2 Gaps: 2

US-09-475-704A-4 (1-1509) x Q6XKD2 (1-495)

QY 1 ATGGGCGCGCGCGCGCATCTGGCGGCGGAGAGCTGGACAACTGGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCGCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysIleTyrMetIleLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGGAGGGCTTCGCGCTGAACCGCGGCTGTGGAGACCGCGGAGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheSerLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnLeu 60  
QY 181 ATGAAGCAGCTGCAAGCGCGCGCTGCAAGACCGGAGAGCTGCGAGCTGTACAAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuPheAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGACGCGCGGATCGAGTCCGCGACACCAAGAGAGCC 300  
Db 81 ThrIleAlaThrLeuTyrCysValHisAlaAspIleAlaValAsnAspThrLysGluAla 100

QY 301 CTGGAACAAGATCGAGAGAGGACGACAACAAGTCCCAAGACCAAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerGlnGlnIleValGlnGlnIle----- 117  
QY 361 GCCGACGGCAAGGTGAGCCAGACTACCCCATGCTGCAGAACTGCGAGGGCCAGATGTG 420  
Db 118 AlaAspLysLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMetVal 137  
QY 421 CACCAAGCCATCAGACCCCGCCAGACCCCTGAGCGCTGGTGAAGGTGATCGAGGAGAGCC 480  
Db 138 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValValGluGluLysAla 157  
QY 481 TTCAGCCCGGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCGCAGGAC 540  
Db 158 PheSerProGluIleIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 177  
QY 541 CTGAACACGATGTTGAACACCGGTGGCGCGCCAGCCCATGCAATGCTGCAAGGAC 600  
Db 178 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 197  
QY 601 ACCATCAACGAGAGAGCGCGCGGAGTGGAGCCGCTGCACCCCGTGCAGGCGCGCCGCTG 660  
Db 198 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValGlnAlaGlyProIle 217  
QY 661 GCCCGCGCGCAGATGCGCGGACCGCGCGGCGGACATCGCGCGCGCCACCAAGCATCCTG 720  
Db 218 AlaProGlyGlnIleArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 237  
QY 721 CAGAGACGATGCGCTGATGATGACCAACCCCGCTGCGCGCTGGCGGACATCTACAAG 780  
Db 238 GlnGluGlnIleThrTrpMetThrAsnAsnProPrometProValGlyAspIleTyrLys 257  
QY 781 CGGTGATCATCTGGCGCTGGAACAAGATGCTGCGGATGATACAGCCCGCTGACATCCTG 840  
Db 258 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 277  
QY 841 GACATCCGCGCAGCGCGCCCAAGAGAGCCCTTCGCGCATACGTTCTTCAAGACC 900  
Db 278 AspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 297  
QY 901 CTGCGCGCGCGCAGGCGCCAGGACCGTGAAGAACTGATGACCGAGACCTGCTGTG 960  
Db 298 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 317  
QY 961 CAGAACGCCAACCCCGCATGCAAGACCATCTCGCGGCTTCGCGCGCGCGCCACCCCTG 1020  
Db 318 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 337  
QY 1021 GAGGAGATGATGACCGCGTGAAGAGGCGGTGGCGCGCGCGCCACCAAGCGCGTGTG 1080  
Db 338 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgValLeu 357  
QY 1081 GCCGAGCGCATGAGCCAGGCGCAACAGCGTGAACATCATGATGCAAGAAGCACTTCAAG 1140  
Db 358 AlaGluAlaMetSerGlnThrAsnSerValAsnIleMetMetGlnArgSerAsnPheLys 377  
QY 1141 GGCGCGCGCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGAGGCGCACATCGCCAAAGAC 1200  
Db 378 GlyProArgLysIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 397  
QY 1201 TGCGCGCGCGCGCGCAAGAGGGGCTGCTGGAAGTGGCGGCAAGAGGGCGCCACGATGAAG 1260  
Db 398 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 417  
QY 1261 GACTGCACCGAGCGCGCAGGCGCAACTTCTGCGGCAAGATCTGGCCAGCCACCAAGGGCGCG 1320  
Db 418 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 437  
QY 1321 CCCGGCAACTTCTGCGAGAACCGCAGCGAGCGCGCGCGCGCCACCGTGGCGCACCGCGCCC 1380  
Db 438 ProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProGluProThrAlaPro 457

QY 1381 CCGCGAGAGCTTCGGCTTGAAGAGACCAACCCCGCCCAAGCAGAGCCCAAGGAC 1440  
Db 458 ProAlaGluSerPheArgPheGluGluThrProAlaProLysGlnGluProLysAsp 477  
QY 1441 CGCGAGCCCTAACCGGAGCCCTGACCGCCCTGCGCAGCCTGTCGGCAGCGGCCCTG 1500  
Db 478 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 493  
QY 1501 AGCCAG 1506  
Db 494 SerGln 495

RESULT 6

Q8UTAT7 PRELIMINARY; PRT; 492 AA.  
ID Q8UTAT7  
AC Q8UTAT7;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21988475; PubMed=11991972;  
RX DOI=10.1128/JVI.76.11.5435-5451.2002;  
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,  
RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,  
RA Foley B.T., GaoJekwe S., Rybak N., Gaseitsiwe S., Vanberg F.,  
RA Marlink R., Lee T.-H., Essex M.,  
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:  
RT consensus sequence for an AIDS vaccine design?";  
RL J. Virol. 76:5435-5451(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,  
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vanberg F.,  
RA Marlink R., Lee T.-H., Essex M.,  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF443095; AAL34742.1; -.  
DR HSSP; P24736; INCP.  
DR GO; GO:0019012; C:virlon; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HC2NFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; ZNF\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.  
SQ SEQUENCE 492 AA; 54948 MW; 1F6343E85D882EB4 CRC64;

Alignment Scores:

Pred. No.: 2.06e-97 Length: 492  
Score: 2440.00 Matches: 464  
Percent Similarity: 94.42% Conservative: 10  
Best Local Similarity: 92.43% Mismatches: 18  
Query Match: 85.92% Indels: 10  
DB: 2 Gaps: 2

US-09-475-704A-4 (1-1509) x Q8UTAT7 (1-492)

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGAGAGCTGGACAAGTGGAGAGAGATCCGC 60  
Db 1 MetGlyAlaSerAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGCGCAGAGCACTACTGCTGAAGCACTGTGTGGGCCAGCGCGAG 120  
Db 1 CTGGCGCGCGCGCGCAGAGCACTACTGCTGAAGCACTGTGTGGGCCAGCGCGAG 120

Db 21 LeuArgProGlyGlyLysLysLysTyraArgLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGGCTTGGCCCTGAAACCCCGGCTGTGAGAGCCGCGAGGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaLeuAsnSerGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60  
QY 181 ATGAGCAGCTGCAGCCCGCCCTGCAGACCCGCGACCGAGGAGCTGCGCAGCTGTACAAC 240  
Db 61 IleYegInLeuGlnProAlaLeuGlnThrGlyThrGluInLeuLysSerLeuTyraAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGCCCGGCGATCGAGGTCCGCGACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrcysValHisAlaGlyIleGluValArgAspThrLysGluAla 100  
QY 301 CTGACAGATCGAGAGAGCAGAACAAAGTCCACGAGAACAGACCAAGCCCAAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysGlu 120  
QY 361 GCCGAGCGCAAGGTGAGCCAGAACTACCCCATCGTGCAAACTTGCAAGGGCCAGATGTTG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTyProIleValGlnAsnLeuGlnGlnMetVal 140  
QY 421 CACCAAGCCATCAGCCCGCGACCCCTGAAACGCTGGGTGAAGTGATGAGAGAGGCC 480  
Db 141 HisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCCCGAGGTGATCCCATGTTCAACCGCCCTGACGAGGCGGCCACCCCGAGAGC 540  
Db 161 PheSerProGluValIlePrometPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCCCGCATGCAATGCTGAAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGGCGCCGAGTGGACCGCTGCAACCCCGTGCCCGTGGCGAGCATCTAAC 660  
Db 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProVal 220  
QY 661 GCCCGCGCGAGATGCGCGAGCCCGCGCGCGAGCAGACATGCGCGCGCCACAGCACCTG 720  
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGAGCAGATCGCGCTGATGACCAACCAACCCCGTGCCCGTGGCGAGCATCTAACAG 780  
Db 241 GlnGluGlnIleThrTrpMetThrSerAsnProProValProValGlyAspIleTyrls 260  
QY 781 CGGTGATCATCTCGGCGCTGAACAAGATCGTGGAGTACAGCCCGGTGAGCATCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrlSerProValSerIleLeu 280  
QY 841 GACATCCGCCAGGGCGCCCAAGAGAGCCCTTCGCGCACTACGTGAGCCGCTTCTCAAGACC 900  
Db 281 AspIleLysGlnGlyProLysGluProPheArgAspTyrlValAspArgPhePheLysVal 300  
QY 901 CTGGCGCGCGAGAGCCCAAGAGAGCTGAGAGACTGATGACCGGAGACCTGCTGTTG 960  
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAGCCCAACCCCGAAGTGAAGACCATCTGCGGCTCTCGGCCCGCGCCACCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProAlaAlaSerLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCGCCCAAGAGCCCGGCTG 1080  
Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyProSerHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGAGCCAGGCCCAACAGCGTGAACATCATGATGACAGAGCACTTCAAG 1140  
Db 361 AlaGluAlaMetSerGlnAlaAsnThrThrAsnIleMetMetGlnLysSerAsnHelys 380  
QY 1141 GGCCCGCGCGCGCAAGCTCAAGTCTTCAACTGCGCGCAAGAGGGCCACATCGCCCAAGAAC 1200  
Db 381 GlyProArgArgThrValLysCysPheAsnCybeLysGluGlyHisIleAlaLysAsn 400

QY 1201 TGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGGCAAGAGGGCCACCATGTAAG 1260  
DB 401 CysArgAlaProArgLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420  
QY 1261 GACTGCACGAGCGCCAGGCGCAACTTCTGGGCAAGATGTGGCCCAAGGAGGCGCGC 1320  
DB 421 AspCysThrGluArgGlnAlaAsnPhenLeuGlyLysIleTrpProSerHisLysGlyArg 440  
QY 1321 CCCGGCAACTTCTGTGAGAACCGCAGCGCCCGCCGCCACCGTGCCCAAGCGCCCGC 1380  
DB 441 ProGluAsnPhenLeuGlnSerArgProGlu-----ProThrAlaPro 454  
QY 1381 CCCGCGCAGAGCTTCCGCTTGTGAGAGACCAACCCCGCCCAAGCAGGAGCCCAAGGAC 1440  
DB 455 ProAlaGluSerPheArgPheGluGlnThrThrProAlaProLysGlnGluProLysAsp 474  
QY 1441 CGCGAGCCCTACCGCGAGCCCTGTACCGCCCTGTGCGAGCGGCGCCCTG 1500  
DB 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
QY 1501 AGCCAG 1506  
DB 491 SerGln 492

RESULT 7  
Q8USV8 PRELIMINARY; PRT; 499 AA.  
AC Q8USV8;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2198475; Pubmed=11991972;  
RX DOI=10.1128/JVI.76.11.5435-5451.2002;  
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,  
RA Williams C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,  
RA Foley B.T., GaoLekwe S., Rybak N., Gaseitsiwe S., Vanberg F.,  
RA Marlink R., Lee T.-H., Essex M.,  
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:  
RT consensus sequence for an AIDS vaccine design?";  
RL J. Virol. 76:5435-5451(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,  
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vanberg F.,  
RA Marlink R., Lee T.-H., Essex M.,  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF443112; AAL34891.1; -  
DR HSSP; P24736; INCP.  
DR GO; GO:0019012; C:viroin; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf\_CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Polypeptide.  
SQ SEQUENCE 499 AA; 55681 MM; 0BFB37A38CA5779E CRC64;

Alignment Scores:  
Pred. No.: 2.16e-97 Length: 499  
Score: 2439.50 Matches: 459  
Percent Similarity: 95.23% Conservative: 20  
Best Local Similarity: 91.25% Mismatches: 19

Query Match: 85.90% Indels: 5  
DB: 2 Gaps: 2  
US-09-475-704A-4 (1-1509) x Q8USV8 (1-499)

QY 1 ATGGGCGCGCGCGCGCGCATCTCTGCGCGCGGAGAGAGCTGAGCAAGTGGAGAGATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGlnLysIleArg 20  
QY 61 CTGGGCGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGCTGTGGGCGCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysCysTyrMetIleLysHisIleIleTrpAlaSerArgGlu 40  
QY 121 CTGGAGGGCTTCCGCTTGAACCCCGCGCTGTGAGAGACCGCGGAGCTGCGAGCTGTACAAC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrAlaGluGlyCysThrGlnIle 60  
QY 181 ATGAAGCAGCTGACAGCCCGCGCTGCAGACCGGCGCATCGAGTTCGCGACCAAGAGGCGC 240  
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuPheAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGGTGCACGCGCGCATCGAGTTCGCGACCAAGAGGCGC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisAlaGlyTleAspValArgAspThrLysGluAla 100  
QY 301 CTGGACAAGATTCAGAGAGAGAGAGAGAGAGATCCAGCAGAGAGAGAGAGAGAGAGAG 360  
DB 101 LeuAspLysIleGluGluGlnLysSerGlnGlnLysThrGlnGlnAlaAla 120  
QY 361 GCCGACGCGCAAGGTGAGCGCGCAACTACCCCATGCTGCAGAACTGCAGGCGCGCATGTGTG 420  
DB 121 AlaAspLysLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACCGAGCCATCAGCCCGCGCAACCTGACGCGCTGGGTGAAGGTGATGAGAGAGAGGCGC 480  
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160  
QY 481 TTCAGCCCGCGAGGTGATCCCATGTTCAACCGCGCTGAGCGAGAGGCGCGCCACCGAGAGC 540  
DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACCATGTTGAACACCGGTGGCGCGCGCGCGCATGTCAGATGCTGAAGAGC 600  
DB 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCGTGCACCGCGTGCAGGCGCGCGCGCG 660  
DB 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCCCGCGCGCGAGATGCGCGAGACCCCGCGCGCGAGCGAGCATCGCGCGCGCGCGCATG 720  
DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGAGCAGATCGCGTGTGATGACCAAGCAACCCCGCGTGCCTGGCGCGCGAGCATCTACAAG 780  
DB 241 GlnGluGlnIleAlaIleTrpMetThrAsnAsnProProValProValGlyAspIleTyrLys 260  
QY 781 CGGTGATCATCTGGGCGCTGAAACAGATCGTGGGATGTACAGCCCGCGTGAAGCATCTG 840  
DB 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGCGAGGCGCGCGCAAGAGCGCTTCCGCGCATCGTGAAGCGCTTCTTCAAGACC 900  
DB 281 AspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheArgThr 300  
QY 901 CTGCGCGCGCGAGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAGCGCGCAACCGCGCATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340

Qy	1021	GAGGAGATGATGATCCGGCTTGCCAGGGGGCTGGGGGGCCCCGGCCCAAGAGGGCGGTGCTG	1080
Db	341	GIUGImetmetThrAlaCysGInGIyValGIyGIyProGIyHISlySalArgValLeu	360
Qy	1081	GCCGAGGCGATGAGCCAGGCGCCAAcAGCGTGAACATCATGATGAGAAGAGCAATTCAAG	1140
Db	361	AlAGIuaIameTserGIaIaAsnaSnIleaSnValmetmetGlnArgGlyAsnPhelys	380
Qy	1141	GGCCCCCGGCGCAACGTCAAGTGTCTCAACTGTGGGCAAGAGGGCCACATCGCCAAAGAAC	1200
Db	381	GIyAlaIaIysArgIleValIyScySpheAbnIcysGIyArgGlnGlyHISileAlaIysAsn	400
Qy	1201	TGCGCGCGCCCCCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGGCCACAGATGAAG	1260
Db	401	CysArgAlaProArgIySlySgIyCysTTrpIyCysGIyLysGlnGlyHISileGlnmetLys	420
Qy	1261	GACTGCACCCGAGCGCCAGGCGCCAACTTCTTGGGCAAGATTTGGCCAGCCACAGGGCGCG	1320
Db	421	AspCysThrGIuArgIaGlnAlaAsnPhelengIyLysIleTTrpProSerHISlySgIyArg	440
Qy	1321	CCCCGCAACTTCTCTGCAGAACCCGCAAGCGCCGCGCCGCCCAACCGTG--CCCAACGCGC	1377
Db	441	ProGIyAsnPhelengIaAsnArgProGIuProThrAlaProAlaGlnProThrAla	460
Qy	1378	CCCCCGCGGAGAGCTTCGGCTTGAGAGAGACCACCCCCGCCCAAGCAGAGAGCCCAAG	1437
Db	461	ProProAlaGlnSerPheArgPheGlnGluThrThrProValProLysGlnGlnLys	480
Qy	1438	GACCGCGAGCCCTTACCGCGAGCCCTTGACCGCCCTGCGAGCGCTGTTGGCAGCGGGCCC	1497
Db	481	Asp-----ArgGlnAlaLeuThrSerLeuLysSerLeuPheGlySerAspPro	496
Qy	1498	CTGAGCCAG 1506	
Db	497	LeuSerGln 499	
RESULT 8			
Q8USYS			
ID	Q8USYS	PRELIMINARY;	PRT; 508 AA.
AC	Q8USYS;		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Gag protein.		
GN	Name=gag;		
OS	Human immunodeficiency virus 1.		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
OX	NCBI_TaxID=11676;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21988475; PubMed=11991972;		
RA	DOI=10.1128/JVI.76.11.5435-5451.2002;		
RA	Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,		
RA	Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,		
RA	Foley B.T., Gaolekwe S., Rybak N., Gaseitsiwe S., Vannberg F.,		
RA	Marlink R., Lee T.-H., Essex M.;		
RT	"Human immunodeficiency virus type 1 subtype C molecular phylogeny:		
RT	consensus sequence for an AIDS vaccine design?";		
RL	J. Virol. 76:5435-5451 (2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,		
RA	Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vannberg F.,		
RA	Marlink R., Lee T.-H., Essex M.;		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF443109; AAL34864.1; -.		
DR	HSSP; P24736; INCP.		
DR	GO; GO:0019012; C:virlon; IEA.		
DR	GO; GO:0003676; F:nucleic acid binding; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		
DR	GO; GO:0016032; P:viral life cycle; IEA.		
DR	Pfam; PF00540; Gag_p17; 1.		
DR	Pfam; PF00607; Gag_p24; 1.		

DR Pfam; PF00098; zF-C2HC; 2.  
DR PRINTS; PR00939; C2HC2NFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; zNF\_C2HC; 2.  
DR PROSITE; PS50158; zF\_C2HC; 2.  
KW AIDS; Core protein; Metal-binding; Polyprotein; zinc; zinc-finger  
SQ SEQUENCE 508 AA; 56799 MW; BFC525AB8CAD688 CRC64;

**Alignment Scores:**

Pred. No.:	2.27e-97	Length:	50
Score:	2439.00	Matches:	44
Percent Similarity:	94.88%	Conservative:	33
Best Local Similarity:	88.39%	Mismatches:	20
Query Match:	85.88%	Indels:	6
DB:	2	Gaps:	1

US-09-475-704A-4 (1-1509) x Q8USY5 (1-508)

OY		1	ATGGCGCCGCGGCAGCATCTTCGCGGCGGAGACTGGACAAGTGGAGAAATCCG	60
Db		1	MetGIAlaArgAlaSerValIleuArgGlyGluValLeuAspThrTrpGluArgIleLys	20
OY		61	CTGCGCCCCGCGGCAAGAAGCACTACATGTCTGAAGCACCTGTGTGGCCAGCCGCGAG	120
Db		21	LeuArgProGlyGlyLysIleHisTyMetLeuLysHisIleValITrpAlaSerArgGlu	40
OY		121	CTGAGAGGCTTGCCCTGAACCCCGCCTGTCTGAGAACCGCCGAGGCTGCAAGCAATC	180
Db		41	LeuGLuArgPheAlaLeuAsnProGlyLeuMetCylThrSerGluGlyCysLysGlnIle	60
OY		181	ATGAAGCAGCTGCAGCCCGCCTGCAGACCSCGACCGAGAGCTGCGCAGCCTGTACAAC	240
Db		61	IleGlnGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuPheAsn	80
OY		241	ACCGTGCACCCCTGTACTGCGGTGCACGCCGCGCATCGAGGTCCGCGCACCAAGAGGCC	300
Db		81	ThrIleAlaThrLeuTyrCysValHisLysGlyIleLysValGlnAspThrLysGluAla	100
OY		301	CTGGACAAGATCGAGGAGAGAGCAACAAGTCCCGACGAGAAGACCCGACGCGCAAGSAG	360
Db		101	LeuAspLysIleGluGluGluGlnAsnLysSerGlnGlnLysThrGlnGlnGlnLys	120
OY		361	-----GCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGACAGAAC	402
Db		121	ThrGlnGlnThrGluAlaAlaAlaGlyLysValSerGlnAsnTyrProIleValGlnAsn	140
OY		403	CTGCAGGCGCCAGATGGTGCAACCAAGCCATCAGCCCCCGCACCCCTGAACGCCCTGGTGAAG	462
Db		141	LeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLys	160
OY		463	GTTGATCGAGGAGAAGGCGCTTCAGCCCCCGAGGTGATCCCCTATTTCACCGCCCTGAGCGAG	522
Db		161	ValValGluGluLysGlyPheAsnProGluValIlePrometPheThrAlaLeuSerGlu	180
OY		523	GCGGCCAACCCTCCAGACCTGAACAAGATGTTGAACAACCGTGGGCGCCACCAAGCGGCC	582
Db		181	GlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAla	200
OY		583	ATGCAGATGCTGAAGGACACCATCAACGAGAGGCGCCGCGAGTGGGACCGCCTGCACCCC	642
Db		201	MetGlnMetLeuLysAspThrIleAsnGluGluAlaAlaGluTrpAspArgValHisPro	220
OY		643	GTCGAGGCGCGGCCGCTGGGCCCCCGCCAGATGCGCGCACACCCCGCGGCGAGCGACATGCC	702
Db		221	ValGlnAlaGlyProValAlaProGlyGlnIleArgGluProArgGlySerAspIleAla	240
OY		703	GCGCGCAACCAGACCCCTGCAGGAGCAGATCGCCTGGATGACCAAGCAACCCCGCGTGC	762
Db		241	GlyThrThrSerAsnLeuGlnGluGlnIleThrTrpMetThrSerAsnProProIlePro	260
OY		763	GTCGGCGACATCTACAAGCGGTGGATCATCTCGGCGCTGAACAAGATCGTGGGATGTAC	822
Db		261	ValGlyGluIleTyrLysArgTrpIleValLeuGlyLeuAsnLysIleValArgMetTyr	280

OY 823 AGCCCGTGAGCATCTGGAGCATCCGCGCAGAGGCCCTTCCGGACTACGTG 882  
Db 281 SerProValSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrVal 300  
OY 883 GACCGCTTCTTCAAGACCCCTGCGCGCGCAGAGCCACCAGACGCTGAAGAACTGGATG 942  
Db 301 AspArgPhePhePheThrLeuArgAlaGlnAlaThrGlnAspValLysAsnTrpMet 320  
OY 943 ACCGAGACCCCTGTGTGTCAGAAAGCCCAACCCCGACTGCAAGACCATCTGCGCGCTCTC 1002  
Db 321 ThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeu 340  
OY 1003 GCGCCCGCGCGCACCCCTGGAGAGATGATGACCCCTGCCAGGGCGTGGCGCGCGCGC 1062  
Db 341 GlyProGlyAlaSerLeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGly 360  
OY 1063 CACAGAGCCCGCGCTGTGCGCGAGCGATGAGCCAGCCCAACAGCGTGAACATCATGATG 1122  
Db 361 HisLysAlaArgValLeuAlaGlnAlaMetSerGlnAlaAsnSerThrAsnIleMetMet 380  
OY 1123 CAGAGAGCAACTTCAAGAGGCGCGCGCGCAACGTCAGATGCTTCAACTGCGCGCAAGAG 1182  
Db 381 GlnArgGlyAsnPheLysGlyProLysArgThrIleLysCysPheAsnCysGlyLysGlu 400  
OY 1183 GGCACATCGCCCAAGAACTGCCGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGGCGCAAG 1242  
Db 401 GlyHisLeuAlaArgAsnCysArgAlaProLysLysLysGlyCysTrpLysCysGlyArg 420  
OY 1243 GAGGCGCCACAGATGAAAGACTGCAACGAGCGCGCGCAACTTCTTGGCAAGATCTGG 1302  
Db 421 GlnGlnHisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrp 440  
OY 1303 CCCAGCCCAAGGCGCGCGCGCGCGCAACTTCTTGCAGAAACCGCAGCGAGCGCGCGCGC 1362  
Db 441 ProSerArgGlyGlyArgProGlyAsnPheLeuGlnAsnArgThrGluProThrAlaPro 460  
OY 1363 ACCGTGCCACCGCGCGCGCGCGCGAGAGCTTCCGCTTGCAGAGACCAACCGCGCGCGC 1422  
Db 461 ProGluProThrAlaProProAlaGlnSerPheArgPheGlnGluThrThrProAlaPro 480  
OY 1423 AAGCAGAGCCCAAGAGCGCGCGAGCCTTACCAGCGCGCGCGCGCGCGCGCGCGCGC 1482  
Db 481 LysGlnGluProLysAspArgGlyProTyrArgGluProLeuIleSerLeuLysSerLeu 500  
OY 1483 TTCGGCAGCGCGCGCGCGCGCGAG 1506  
Db 501 PheGlySerAspProLeuSerGln 508  
RESULT 9  
06S7X9 PRELIMINARY; PRT; 497 AA.  
AC 06S7X9;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RG HIV HLA epitope mapping;  
RA Korber B.T., Mullins J.I., Goulder P.J., Brander C., Klepiela P.,  
RA Moore S., Shindo N., Walker B.,  
RU Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY463231; AAR27722.1; -.  
DR HSSP; P12493; 1AFV.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.

DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
DR AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.  
SQ SEQUENCE 497 AA; 55385 MW; 0CD9B58C540552BC CRC64;  
Alignment Scores:  
Pred. No.: 2,39e-97 Length: 497  
Score: 2438.50 Matches: 460  
Percent Similarity: 95.02% Conservative: 17  
Best Local Similarity: 91.63% Mismatches: 20  
Query Match: 85.86% Indels: 5  
DB: 2 Gaps: 2  
US-09-475-704A-4 (1-1509) x Q6S7X9 (1-497)  
OY 1 ATGGCGCGCGCGCGCGCATCTGCGCGCGCGAGAACTGGAAGTGGAGAAATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20  
OY 61 CTGCGCGCGCGCGCGCAAGAAAGCACTACATGCTGAAGCACTGTGTGGCGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysCysTyrMetLeuLysHisIleIleTrpAlaSerLysGlu 40  
OY 121 CTGAGGGCTTCCCTGAAACCCCGCGCTGTGAGACCGCGCGAGGGCTTGAAGCAGATC 180  
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysThrGlnIle 60  
OY 181 ATGAAGCAGCTGCAGCCCGCGCTGCAGACCGCGCACCGAGAGCTGCGACCTGTACAAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGluLeuArgSerLeuTyrAsn 80  
OY 241 ACCGTGGCCACCTGTACTGCGTGACCGCGCGCATCGAGTCCGCGCGCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnLysIleAspValArgAspThrLysGluAla 100  
OY 301 CTGACACAGATCGAGGAGGAGCAGAACTACCCCATCGTGCAAACTGACGGCCAGAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnAsnLysSerGlnGlnLysAlaGlnGlnAlaLysAla 120  
OY 361 GCCGAGCGCAAGGTGAGCGAGCAACTACCCCATCGTGCAAACTGACGGCCAGATGTG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMetVal 140  
OY 421 CACCAAGGCATCAGCCCGCGCAACCTGACGCGCTGGTGAAGGTGATCGAGAGAGGCC 480  
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGluLysAla 160  
OY 481 TTCAGCCCGAGGTGATCCCATGTTCAACCGCCCTGACGAGGGCGCGCAACCCCGAGGAC 540  
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGlnGluAlaThrProGlnAsp 180  
OY 541 CTGAACAGATGTTGAACACCGTGGCGCGCACAGGCGCGCATGACATGCTGAAGAGC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
OY 601 ACCATCAACGAGGAGCGCGCGAGTGGACCGCTGCAACCCCGTGACAGCGCGCGCGTG 660  
Db 201 ThrIleAsnGlnGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
OY 661 GCGCGCGCGCGAGATGCGCGAGCCCGCGCGCGAGCAGCATGCGCGCGCGCGCACCCCTG 720  
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlnGlyThrThrSerThrLeu 240

QY 721 CAGAGCAGATCGCTGTGATGACCAGCAACCCCGCTGCCGTGGGCAATCTACAAG 780  
Db 241 GlnGlnGlnIleAlaTrpMetThrGlyAsnProProValProValGlyAspIleTyrIys 260  
QY 781 CGGTGGATCATCTGGGCTGAAACAAGATCGTGGATGTACAGCCCGGTGAGCATCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGCAGGAGCCCAAGAGAGCCCTCCGCGACTACGTGAGACCCGCTTCTCAAGAC 900  
Db 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheIleThr 300  
QY 901 CTGCGCGCGAGCAGAGCCACCCAGAGACGTGAAGACTGTGATGACCGAGACCCCTGCTGTG 960  
Db 301 LeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAGCGCCAAACCCGACTGCAAGACATCTCGCGCTCTCGCGCCCGCCGACCCCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysIleThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 GAGAGATGATGACCGCTGCGCAGGCGTGGGGCGCCCGCCACAGAGCCCGCTGTG 1080  
Db 341 GlnGlnMetMetThrAlaCysGlnGlyValGlyIleProGlyHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGAGCGCCAGCCACAGCGGTGAACATCATGATGCAAGAGCAACTTCAAG 1140  
Db 361 AlaGlnAlaMetSerGlnAlaAsnSer--AsnIleMetMetGlnArgGlyAsnPheLys 379  
QY 1141 GGGCCCCGGCGCAACGTCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATGCCAAGAAC 1200  
Db 380 GlySerLysArgIleValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaLysAsn 399  
QY 1201 TGCCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACAGATGAAG 1260  
Db 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisGlnMetLys 419  
QY 1261 GACTGCACCGAGCGCCAGCCCAACTTCTTGGCAAGATCTGGCCAGCCCAAGAGGGCGC 1320  
Db 420 AspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCGCGCAACTTCTCTGCAAGAACCGCAGGAGACCGCGCGCCCGCCACCGTGCACCGCGCC 1380  
Db 440 ProGlyAsnPheIleGlnSerArgProGluProThrAlaProLeuGlnProThrAlaPro 459  
QY 1381 CCGCGCGAGAGCTTCCGCTTGCAGAGAACCCCGCGCCCGCCCAAGAGAGAGCCCAAGAG 1440  
Db 460 ProAlaGlnSerPheArgPheGlyGlnThrThrProAlaProLysGlnGlnProGlnAsp 479  
QY 1441 CGGAGACCTTACCGGAGCGCCCTGACCGCCCTGCGAGCCTGTTCCGAGCGGCGCCCTG 1500  
Db 480 -----ArgGluProLeuIleSerLeuLysSerLeuPheGlySerAspProLeu 495  
QY 1501 AGCCAG 1506  
Db 496 SerGln 497  
RESULT 10  
Q90ME1 PRELIMINARY; PRT; 492 AA.  
ID Q90ME1; AC Q90ME1; DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21565559; PubMed=11709098; DOI=10.1089/0892220152644241;  
RA Engelbrecht S., de Villiers T., Sampson C.C., zur Megede J.,

RA Barnett S.W., van Rensburg E.J.;  
RT "Genetic analysis of the complete gag and env genes of HIV type 1  
RT subtype C primary isolates from South Africa.";  
RL AIDS Res. Hum. Retroviruses 17:1533-1547(2001).  
DR EMBL; AF391270; AAK7531.1; -.  
DR HSSP; P24736; INCP.  
DR GO; GO:0019012; C:viralion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polyprotein; Zinc; Zinc-finger.  
SQ SEQUENCE 492 AA; 5499 MW; B74F6064D506635D CRC64;  
Alignment Scores:  
pred. No.: 2.77e-97 length: 492  
Score: 2437.00 Matches: 457  
Percent Similarity: 94.82% Conservative: 19  
Best Local Similarity: 91.04% Mismatches: 16  
Query Match: 85.81% Indels: 10  
DB: 2 Gaps: 2  
US-09-475-704A-4 (1-1509) x Q90ME1 (1-492)  
QY 1 ATGGCGCGCGCGCGCAGCATCTCGCGCGCGGAGAGCTGACAAAGTGGAGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuGlyGlyGlyLysLeuAspThrTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGCGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisIleTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGGGCTTGCCTTGAACCCCGCGCTGTGAGAACCGCGAGGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGCAGCGCGCGCTGCAGACCGGCAACGAGAGAGTCCGCAACCTGTACAAC 240  
Db 61 IleArgGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuLysSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGCGCGCATCGAGGTCCCGCAACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisAlaLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCCAAGACCAAGAGCCCAAGAGAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnGlnLysLysCysGlnGlnLysThrGlnGlnAlaLysGlu 120  
QY 361 GCCGACGGCAAGTGAAGCCAGAACTACCCCATCGTGCAGAACTTGCAGGGCCAGATGTG 420  
Db 121 AlaAspGlyLysIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLysAla 140  
QY 421 CACCAAGCCATCAGCCCCCGCGCAACCTGAAAGCGCTGGAAGGTGATCGAGAGAGAGCC 480  
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLysAla 160  
QY 481 TTCAGCCCCGAGGTGATCCCGCATGTTCAACCGCCCTGAGCGAGGGCCACCCCGCAGAG 540  
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGCGGCGCCACAGCGCCCATGCAATGCTGAAGAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGCACCCCGTGCAGAGCGCGCGCGTG 660  
Db 201 ThrIleAsnGlnGlnAlaAlaGlnTrpAspArgIleHisProValHisAlaGlyProIle 220

```
QY 661 GCCCGCCGAGATGCGGACCCCGCGGACGACATCGCCGCGCCACGACCCCTG 720
Db 221 AlaProGlyGlnMetArgLupProArgGlySerAspIleAlaGlyThrThrSerThleu 240
QY 721 CAGGAGCAGATCGCTGTGATGACCAACACCCCGCTGCGCGGACATCTACAG 780
Db 241 GlnGlnGlnIleAlaIleThrMetThrSerAsnProProValProValGlyGluIleYrlys 260
QY 781 CGGTGATCATCTCGGCTGACACAGATCGGTGATGTACAGCCCGGTGAGCATCTG 840
Db 261 ArgTrpIleIleLeuGlyLeuAsnIleValArgMetTyrSerProValSerIleleu 280
QY 841 GACATCCGCGAGGCGCCCAAGAGAGCCCTCCGCGATACGTGACCCGCTTCTACAGAC 900
Db 281 AspIleIysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheIysThr 300
QY 901 CTGCGCGCGGAGAGGCGCCACGACGATGAGAACTGATGACCGGACCCCTGCTGTG 960
Db 301 LeuArgAlaGlnGlnAlaThrGlnAspValIysAsnTrpMetThrAspThrLeuVal 320
QY 961 CAGAACGCGCAACCCGACTGCAAGACATCTCGCGCTCTCGCGCCGCGCCACCCCTG 1020
Db 321 GlnAsnAlaAsnProAspCysIleIleLeuArgGlyLeuGlyProGlyAlaThrleu 340
QY 1021 GAGGAGATGATGACGCGCTGCGGCGGTGGCGCGCGCCGCGCAAGAGCGCGCTGTG 1080
Db 341 GlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisIleValAlaArgValleu 360
QY 1081 GCCGAGCGATGAGCGCGCAAGCGCTGAACTGATGATGACGAGGCGCCACGACTTCAAG 1140
Db 361 AlaGlnAlaMetSerGlnAlaAsnSerThrAsnIleMetMetGlnArgGlyAsnIleIys 380
QY 1141 GGCGCGCGCGCGCAAGCTCAAGTGTCTCACTGCGGCAAGAGGCGCAACATCGCCAGAAG 1200
Db 381 GlyProLysArgAsnIleIysCysPheAsnGlyIleGlyIleIleValAlaArgAsn 400
QY 1201 TGCGCGCGCGCGCGCAAGAGGCTGTGAACTGCGGCGCAAGAGGCGCCACGACTGAAG 1260
Db 401 CysArgAlaProArgGlyIleGlyCysTrpIleGlyIleGlyIleGlyIleGlyIleMetIys 420
QY 1261 GACTGCAACGAGCGCGCGCAAGCTCTCTGCGCAAGATCTGCGCGCGCGCGCGCGCGCG 1320
Db 421 AspCysThrGlnArgGlnAlaAsnIleGlyIleIleTrpProSerHisIleGlyArg 440
QY 1321 CCGCGCAACTCTCTGCAAGACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 441 ProGlyAsnIleGlnIleAsnArgThrGln-----ProThrAlaPro 454
QY 1381 CCGCGCGAGAGCTTCCGCTTCAAGAGACACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
Db 455 ProAlaGlnSerPheArgPheGlnGlnThrAsnProAlaProLysGlnGlnIleProLysAsp 474
QY 1441 CGGAGACCTTACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
Db 475 -----ArgGluProLeuThrSerLeuIysSerLeuPheGlySerAspProSer 490
QY 1501 AGCCAG 1506
Db 491 SerGln 492
RESULT 11
Q9QN98 PRELIMINARY; PRT; 491 AA.
AC Q9QN98;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Gag.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=93IN101;
RX MEDLINE=99433500; PubMed=10505681; DOI=10.1089/089222299310223;
RA Mochizuki N., Otsuka N., Matsuo K., Shino T., Kojima A., Kurata T.,
RA Sakai K., Yamamoto N., Isomura S., Dhole T.N., Takebe Y., Matsuda M.,
RA Tatsumi M.;
RT "An infectious DNA clone of HIV type 1 subtype C.";
RL AIDS Res. Hum. Retroviruses 15:1321-1324(1999).
DR EMBL; AB023804; BAA85225.1; -.
DR HSSP; P24736; INCP.
DR GO; GO:0019012; C:viralion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR010999; Retrovir_matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIVMATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS0158; ZF_CCHC; 2.
KW AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.
SQ SEQUENCE 491 AA; 54809 MW; 0B719EF5C77148B CRC64;
```

```
Alignment Scores:
Pred. No.: 3.22e-97 Length: 491
Score: 2435.50 Matches: 463
Percent Similarity: 95.42% Conservative: 16
Best Local Similarity: 92.23% Mismatches: 12
Query Match: 85.76% Indels: 11
DB: Gaps: 3
```

US-09-475-704A-4 (1-1509) x Q9QN98 (1-491)

```
QY 1 ATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyIleValAspLysTrpGluArgIleArg 20
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 21 LeuArgProGlyGlyIleLysIleIleIleIleIleIleIleIleIleIleIleIleIle 40
QY 121 CTGAGGCGCTTCCGCTTGAACCGCGCGCTGTGAGACCGCGCGCGCGCGCGCGCGCGCG 180
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnIleValAlaGlnGlyCysIleGlnIle 60
QY 181 ATGAAGCAGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 61 IleLysGlnLeuGlnIleProAlaLeuGlnIleThrGlyThrGlnGlnIleValAsn 80
QY 241 ACCGTGCGCACCTGTACTGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGlnValArgAspThrIleGlnAla 100
QY 301 CTGACACAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
Db 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY 361 GCCGACGCGCAAGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 121 AlaAspGlyLysValSerGlnAsnIleProIleValGlnAsnLeuGlnGlnGlnMetVal 140
QY 421 CACCGCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaIleValIleValIleValIleValIle 160
QY 481 TTCAGCGCGCGAGGTGATCCCATGTTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
```

Db 161 PheSerProGluValIleProMetThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGCGGCCACAGCCCGCATGATGCTGAAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlnIleAlaIleMetGlnMetLeuAsp 200  
QY 601 ACCATCAACGAGAGCGCCGCGAGTGGGACCGCTGCACCCCGTGACGCGCGCCGTG 660  
Db 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProIleHisAlaGlyProIle 220  
QY 661 GCCCGCGCGCAGATGCGCGACCCCGCGCGCAGCAGATCGCCGCGCCACAGACCCCTG 720  
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerSerLeu 240  
QY 721 CAGGAGCAGATCGCCTGGATGACCAACACCCCGCTGCGCCGCGGACATCTTACAAG 780  
Db 241 GlnGluGlnIleAlaThrMetThrGlyAsnProProValProValGlyAspIleTryLys 260  
QY 781 CGGTGGATCATCTGGGCTGGAACAAGATCGTGGGATGTACAGCCCGTGAGCATCCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280  
QY 841 GACATCCGCGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTTCTCAAGAC 900  
Db 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLysThr 300  
QY 901 CTGCGCGCGCAGCAGCCACCCAGACGTGAAGAACTGGATGACCGAGACCCCTGCTGTG 960  
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320  
QY 961 CAGAAACGCCAACCCCGACTGCAAGAACATCTCGCGGCTCTCGCGCCCGCGCCACCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGGCGGCGCCCGGCAAGAGCCCGCTGCTG 1080  
Db 341 GlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360  
QY 1081 GCCGAGCGGATGAGCCAGGCGCAACACCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140  
Db 361 AlaGluAlaMetSerGlnAlaAsnSer--ThrIleLeuMetGlnArgSerAsnLys 379  
QY 1141 GGGCGCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGCGCACATCGCCAAAGAAC 1200  
Db 380 GlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlnHisIleAlaLysAsn 399  
QY 1201 TGCCGCGCGCCCGCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCCACAGTGAAG 1260  
Db 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisGlnMetLys 419  
QY 1261 GACTGCACCGAGCGCCAGGCGCAACTTCTTGCGCAAGATCTGCGCAGCCACAAAGGCGCG 1320  
Db 420 AspCysThrGlnArgGlnAlaAsnLysLeuGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCGCGCAACTTCTGCAAGAACCGCAGAGCCCGCGCCCGCCACCGTCCCGCAGCGCCCC 1380  
Db 440 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 453  
QY 1381 CCGCGCGAGAGCTTCGCTTCGAGAGAGACCAACCCCGCCCGCCAGAGAGAGCCCAAGAG 1440  
Db 454 ProAlaGlnSerPheArgPheGlnGluThrThrProAlaProLysGlnGluProLysAsp 473  
QY 1441 CGGAGACCCCTACCGCAGAGCCCGCTGACCGCCCTGCGCAGCCCTGTCGCGAGCGCC 1500  
Db 474 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspLeuLeu 489  
QY 1501 AGCCAG 1506  
Db 490 SerGln 491  
RESULT 12  
Q901X0

ID Q901X0 PRELIMINARY; PRT; 492 AA.  
AC Q901X0;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21565558; PubMed=11709097; DOI=10.1089/08892220152644232;  
RA van Harmelen J., Williamson C., Kim B., Morris L., Carr J.,  
RA Abdool Karim S.S., McCutchan F.;  
RT "Characterization of full-length HIV type 1 subtype C sequences from  
RT South Africa."  
RL AIDS Res. Hum. Retroviruses 17:1527-1531(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA van Harmelen J.H., Williamson C., Kim B., Morris L., Carr J.,  
RA Abdool Karim S.S., McCutchan F.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY043176; AAL05334.1; -.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF0098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; ZnF\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polypotein; Zinc; Zinc-finger.  
SQ SEQUENCE 492 AA; 55075 MW; 11966A2F3D272ED0 CRC64;  
  
Alignment Scores:  
Pred. No.: 4,12e-97 Length: 492  
Score: 2433.00 Matches: 461  
Percent Similarity: 94.62% Conservative: 14  
Best Local Similarity: 91.83% Mismatches: 17  
Query Match: 85.67% Indels: 10  
DB: 2 Gaps: 2  
  
US-09-475-704A-4 (1-1509) x Q901X0 (1-492)  
QY 1 ATGGCGCGCGCGCCAGCATCTGCGCGCGAGAACTGGACAACTGGAGAAATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCAAGAACTACATGCTGAAGCACTGCTGTGGCGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysCysTyrMetLeuLysHisIleIleTrpAlaSerArgGlu 40  
QY 121 CTGAGGCGCTTCGCCCTGAACCCCGCGCTGCTGAGACCGCGAGGCGCTGCAAGATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrLysGluGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGCAGCCCGCGCTGCAGACCGGCAACGAGAGCTGCGAGCCTGTACAAC 240  
Db 61 IleAsnGlnLeuHisProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCTGTCACGCGCGCATGAGGTCCGACACCAAGAGAGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGluIleGluValArgAspThrLysGluAla 100  
QY 301 CTGACAAGATGAGAGAGAGAGAGAACTCCAGAGAAAGCCAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnAsnIleGlnIleLysThrGlnGlnAlaLysAla 120

```
QY 361 GCCGACGGCAAGTGAGCCGAAGTACCCCATCGTGCAAGAACTGACAGGCCAGATGCTG 420
Db 121 Alaasgilylvalserglnasntrproilevalglnasnleuglnglylmetval 140
QY 421 CACGAGGCCATCAGCCCCCGCAACCTGAACGCTGGGTGAAGTGATGAGAGAGGCC 480
Db 141 Hisglnalalieserproargthrleuasnalatrvallyvalileglnlulysala 160
QY 481 TTCAGCCCCGAGGTATCCCCATGTTCAACCGCCCTGAGCGAGGGCGGCCACCCCGAGAC 540
Db 161 PheSerProgluvalilePrometPheThrAlaleuSerGluGlyAlaThrProglinaSp 180
QY 541 CTGAACACGATGTTGAACACCGTGGCGGCCACAGCGGCCATGACAGATGCTGAAGGAC 600
Db 181 LeuasnThrmetLeuasnThrValGlyGlyHisglnalalameGlnmetLeuylsasp 200
QY 601 ACCATCAACGAGGAGCGCCGAGTGGGACCGCCTGCACCCCGTGCAAGCGCGCGCCGCTG 660
Db 201 ThrileasnGlnGluAlaGluTTPaspArgLeuHisProvalHisAlaGlyProval 220
QY 661 GCCCGCCGCGCAGATGCGCGCAGACCCCGCGCGCAGCAGATCGCGCGCGCCACGACCCCTG 720
Db 221 AlaProGlyGlnmetArgGluProArgGlySeraspIleAlaGlyThrThSerThrleu 240
QY 721 CAGGACGAGATCGCCTGGATGACCAACACCCCGCGTGCCTGGCGCGCAGACATCTACAAG 780
Db 241 GlnGlnGlnIleAlaTrpMetThrAlaasnProIleProvalGlyaspIleTyrLys 260
QY 781 CGGTGATCATCCTGGGCTTGAACAAGATCGTGGGATGTACAGCGCCGTGAGCATCCTG 840
Db 261 ArgTrpIleIleleuGlyLeuasnLysIleValArgMetTyrSerProvalSerileu 280
QY 841 GACATCCGCGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGAC 900
Db 281 AspIleArgGlnGlyProLysGluProPheArgaspTyrValaspArgPhePheLysThr 300
QY 901 CTGCGCGCGCAGAGCGCCACCAAGAGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGAC 960
Db 301 LeuArgAlaGlnGlnAlaThrGlnaspVallylaspantTrpMetThrGluThrleuVal 320
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCCTGGCGCTCTCGGCGCCCGCGCGCACCTG 1020
Db 321 GlnasnAlaasnProaspCysLysThrIleLeuArgAlaleuGlyProGlyAlaSerleu 340
QY 1021 GAGGAGATGATGACCGCCTGCGAGGGCGTGGCGCGCCCGCGCCACAGGCGCGTGTG 1080
Db 341 GluGlnmetmetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValleu 360
QY 1081 GCCGAGGCGATGAGCCAGGCCCAACAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140
Db 361 AlaGlnAlaMetSerGlnAlaasnasnIleasnIleMetMetGlnLysSerAsnPhelys 380
QY 1141 GGGCCCCGCGCAAGTCAAGTCTTCAACTGCGGGCAAGAGGGCCACATGCGCAAGAAC 1200
Db 381 GlyProLysArgIleValLysCysPheasnCysGlyLysGlnGlyHisIleAlaArgasn 400
QY 1201 TGCCGCGCGCCCGCGCAAGAGGCTGTGAAGTGGCGCAAGAGGGCCACAGATGAAG 1260
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisGlnmetLys 420
QY 1261 GACTGACCGAGCGCCAGGCCCACTTCTGGGCAAGATCTGGCCACGACCAAGGGCGCG 1320
Db 421 AspCysIleGluArgGlnAlaasnPhelenglyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGCGCAACTTCTGCAAGACCGCAGCGAGCGCGCGCGCCCGCCACCGTGGCCACCGCCCC 1380
Db 441 ProGlyAsnPhelenglnasnArgProGlu-----ProThrAlaPro 454
QY 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGACACCCCGCGCCCGCCACAGAGAGCCCAAGAC 1440
Db 455 ProAlaGlnSerPheLysPheGlnGluGluThrThrProAlaProLysGlnGluThrLysasp 474
QY 1441 CGCGAGCCCTAACCGGAGCGCCCTGACCGCCCTGCGCAGCCTGTCGCGACGGCGCCCTG 1500
```

```
Db 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490
QY 1501 AGCCAG 1506
Db 491 SerGln 492
```

## RESULT 13

```
Q7SLX7 PRELIMINARY; PRT; 491 AA.
ID Q7SLX7
AC Q7SLX7;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Gag protein.
GN Name=gag;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RA Gupta R.M., Prasad V., Singh B., Seth P.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF533131; AAP87971.1; -.
DR HSSP; Q70622; 1HVN.
```

```
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR010999; Retrovir matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS0158; ZF_CCHC; 2.
KW AIDS; Core protein; Polyprotein.
SQ SEQUENCE 491 AA; 54815 MW; 36CD9A56CAEFAD7C CRC64;
```

## Alignment Scores:

```
Pred. No.: 5.28e-97 Length: 491
Score: 2430.50 Matches: 463
Percent Similarity: 94.82% Conservative: 13
Best Local Similarity: 92.23% Mismatches: 15
Query Match: 85.58% Indels: 11
DB: 2 Gaps: 3
```

US-09-475-704A-4 (1-1509) x Q7SLX7 (1-491)

```
QY 1 ATGGGGCGCCCGCGCCAGCATCCTGCGCGCGCGAGAACTGGACAAGTGGAGAAATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGlnLysIleArg 20
QY 61 CTGCGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCAACCTGCTGGCGCGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGln 40
QY 121 CTGAGGGCTTCCGCTTGAACCCCGCGCTGTGAGAGACCGCGAGGGCTGCAAGCAAGTC 180
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysLysGlnIle 60
QY 181 ATGAAGCAGCTGACGCGCCCGCGCTGACAGCCGCGACCGAGAGTGGCGAGCGCTGACAC 240
Db 61 LeuLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGCTGACAGCGCGCGCATCGAGGTCCGCGACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGlnValArgAspThrLysGlnAla 100
```

QY 301 CTGACAAGATCGAGAGAGACGAACAAGTCCACGACAGAAGACCCAGGACGAG 360  
Db 101 Leuapblysiileglugluglunabnlysglyglnglnlythrnglnalalyggly 120  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGACGAACCTGACGGGCCAGATG 420  
Db 121 Alaapbglylyvalserglnabnlyrproilevalglnabnleuglnglylmetval 140  
QY 421 CACCAAGCCATCAGCCCCCGCACCCCTGAACGCGCTGGTGAAAGTGATGAGGAAGGCC 480  
Db 141 Hisglnalalieserproargthrleuabnlatrpvallylsvalilegluglulysala 160  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGGCCACCCAGGAC 540  
Db 161 Pheserprogluvalilleprometpnethratalaenusergluglyalatrproglinasp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGGGGCCACCAAGCGCCCATGACATGCTGAAGAC 600  
Db 181 leuasnthrmetleuasnthrvalglyglyhisglnalalametglnmetleuysasp 200  
QY 601 ACCATCAACGAGAGGCGCCGAGTGGGACCGCCTGCAACCCGTCGACGCGCGCCGCTG 660  
Db 201 Thrilleasnugluglualaalaglutrpspargleuhisprovalhisalaglproile 220  
QY 661 GCCCGCCGCGCAGATGCGCGACCCCGCGGCGACGACATCGCGCGGCCACCAAGCCTG 720  
Db 221 Alaaprogluglmetearggluproargglyseraspilleagllythrthrsertleu 240  
QY 721 CAGAGCAGATCGCCTGATGACCAACCAACCCCGCTGCGCGTGCGACATCTACAAG 780  
Db 241 Glnuglunilleatrpmetthrashasnproprovalprovalglyaspilleyrlys 260  
QY 781 CCGTGATCATCTGGGCTGAAACAAGATCGTGCGATGTACAGCCCGTGAGCATCTG 840  
Db 261 Argtrpilelleleuglyleuabnlysilvalargmettyrserprovalserileu 280  
QY 841 GACATCCGCGAGGCGCCCAAGAGAGCCCTCCGCACTACGTGACCGCTTCTTCAAGAC 900  
Db 281 Aspilleysglnuglyprolysglupropheargaspryvalasprgphelysthr 300  
QY 901 CTGCGCGCGAGAGGCGCCACCAAGACGTGAAGACTGATGACCGACCCCTGCTG 960  
Db 301 LeuarglialuglinalaThrlnabrvallylsasntrpmetthrapsrthrleuval 320  
QY 961 CAGAACGCCAACCCGACTGCAAGACATCTCGCGCTCTCGCGCGCGGCCACCCCTG 1020  
Db 321 GlnabnalasnproaspCyslysthrilleuargalaleuglyproglyalatrleu 340  
QY 1021 GAGGAGATGATGACCGCCTGACAGGGCGTGGCGCGCCGCGCACAGGCGCGCTG 1080  
Db 341 GluglumetmetthrAlaCysglnlyValglyProserhielysAlaArgvalleu 360  
QY 1081 GCCGAGGCGATGAGCGGCAAGGCGGTGAACATCATGATGACAGAAGCAACTTCAAG 1140  
Db 361 AlagluAlametserglnThrlyser---ThrilemetmetGlnargSerAsnPhelys 379  
QY 1141 GGCCCCCGCGCAAGTCAAGTGTCTCAACTGCGGCAAGAGGGGCCACATCGCCAAGAC 1200  
Db 380 GlySerlysarglilevallyscyphesasnCyseglylysgluglyhisileAlaArgasn 399  
QY 1201 TGCCGCGCGCGCGCAAGAAGGCTGCTGAAGTGGCGCAAGAGGGCCACCATGAAG 1260  
Db 400 CyseargAlaProarglylysglyCystrplysCyseglylysgluglyhisileGlnmetlys 419  
QY 1261 GACTGCAACCGAGCGCGCAACTTCTGGGCAAGATCTGGCCCAAGAGGGCGCGC 1320  
Db 420 AspCystrgluargglnAlaasnPheluglylysriletrProserhislyseglyArg 439  
QY 1321 CCCGGAACCTTCTGAGAACCGGACGAGCGCGCGCCCGCCACCGTGCACCGCGCCC 1380  
Db 440 ProglyAsnPheluglnSerArgProglu-----ProthralaPro 453  
QY 1381 CCCGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGCCCGCAAGAGAGCGCCAAGAC 1440

Db 454 ProAlaGluSerPheargPheglugluThrThrProAlaProlysglngluLeuysAsp 473  
QY 1441 CGCGAGCCCTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTCGCGAGCGCCCTG 1500  
Db 474 -----ArggluProleuThrSerLeuysSerLeuPhegllySerAspProleu 489  
QY 1501 AGCCAG 1506  
Db 490 SerGln 491  
RESULT 14  
Q901Z7 PRELIMINARY; PRT; 492 AA.  
ID Q901Z7 AC Q901Z7; DT 01-DEC-2001 (TREMBLrel. 19, Created) DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update) DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update) DE Gag protein. GN Name=gag; OS Human immunodeficiency virus 1. OC Viruses; Retroid viruses; Retroviridae; Lentivirus. ON NCBI\_TaxId=11676; [1] SEQUENCE FROM N.A. RP MEDLINE=2156558; PubMed=11709097; DOI=10.1089/08892220152644232; RX van Harmelen J., Williamson C., Kim B., Morris L., Carr J., RA Abdoel Karim S.S., McCutchan F.; RT "Characterization of full-length HIV type 1 subtype C sequences from RT South Africa."; RL AIDS Res. Hum. Retroviruses 17:1527-1531 (2001). [2] SEQUENCE FROM N.A. RP van Harmelen J.H., Williamson C., Kim B., Morris L., Carr J., RA Abdoel Karim S.S., McCutchan F.; RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases. DR EMBL; AY043173; AAL05307.1; -. DR HSSP; P24736; INCP. DR GO; GO:0019012; C:vitrion; IEA. DR GO; GO:0003676; F:nucleic acid binding; IEA. DR GO; GO:0005198; F:structural molecule activity; IEA. DR GO; GO:0016032; P:viral life cycle; IEA. DR Pfam; PF00540; Gag\_p17; 1. DR Pfam; PF00607; Gag\_p24; 1. DR Pfam; PF00098; zf\_CCHC; 2. DR SMART; SM00343; ZnF\_C2HC; 2. DR PROSITE; PS50158; zf\_CCHC; 2. KW AIDS; Core protein; Polyprotein. SQ SEQUENCE 492 AA; 55205 MW; 55CE9EB2D591BA3A CRC64;  
Alignment Scores:  
Pred. No.: 5.55e-97 Length: 492  
Score: 2430.00 Matches: 458  
Percent Similarity: 94.82% Conservative: 18  
Best Local Similarity: 91.24% Mismatches: 16  
Query Match: 85.56% Indels: 10  
DB: 2 Gaps: 2  
US-09-475-704A-4 (1-1509) x Q901Z7 (1-492)  
QY 1 ATGGCGCGCGCGCCAGCATCTCTGCGCGCGGCAAGACTGACAAGTGGAGAAGATCCGC 60  
Db 1 MetcglYAlaargAlaSerileuArgglyglulysleuAspThrtrpGlulysrileArg 20  
QY 61 CTCGCGCGCGCGCGCAAGACCACTACATGCTGAAGCAACCTGCTGGGCCAGCGCGAG 120  
Db 21 LeuArgProgllyglylylysbYstrymetleuysHisileValTrpAlaSerArgglu 40  
QY 121 CTGAGAGGCTTCCCTGAGAACCGCGCGCGCTGCTGGAAGACCGCGGAGGCTGCAAGCATC 180  
Db 41 LeugluargPheSerLeuasnProglyLeuengluThrserGlnuglyCylysglnille 60  
QY 181 ATGAAGCAGCTGAGCGCGCGCTGCAAGACCGGCAACGAGAGAGCTGCGAGCTGTACAAC 240

```
Db      61 MetLysGlnLeuGlnProAlaLeuGlnThrGlnGluLeuLysSerLeuTyrAsn 80
Qy      241 ACCGTGGCCACCTGTACTGCGTGACAGCCGGCATCGAGTCCGGCACCAAGAGGCC 300
Db      81 ThrValAlaThrLeuPheCysValHisGlnLysIleAlaValArgAspThrLysGlnAla 100
Qy      301 CTGACACAGATCGAGAGGAGCAGAAAGTCCAGCAGAGAAGACCCAGAGGCCAAGAG 360
Db      101 LeuAspLysIleGlnGlnGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120
Qy      361 GCCGAGCGCAAGGTGAGCCAGAACTACCTCATCGTGAGAACCTGAGGGCCAGATGGTG 420
Db      121 AlaAspGlyThrValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMetVal 140
Qy      421 CACCAAGCCATCAGCCCCCGACACCTGTAAGCCTGGGTGAAGGTGATCGAGAGAGGCC 480
Db      141 HisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValIleGlnLysAla 160
Qy      481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCGCAGAC 540
Db      161 PheSerProGlnValIlePheMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 180
Qy      541 CTGAACACGATGTTGAACACCGTGGGGCGCCACAGCCCGCATGAGATGCTGAAGAC 600
Db      181 LeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLysAsp 200
Qy      601 ACCATCAACGAGAGGCGCCGAGTGGGACCGCCTGCAACCCCGTGAGCGCCGCCCTG 660
Db      201 ThrIleAsnGlnGlnAlaGlnAlaGlnTyrAspArgLeuHisProValHisAlaGlyProIle 220
Qy      661 GCCCGCGCGCAGATGCGCGACCCCGCGCAGCGACATCGCCGCGCCACAGACCCCTG 720
Db      221 AlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240
Qy      721 CAGAGCAGATCGCTGTGATGACCAACCCCGCTGCGCCGTGGCGGACATCTTCAAG 780
Db      241 GlnGlnGlnIleAlaTyrMetThrAsnAsnProProValProValGlyAspIleTyrLys 260
Qy      781 CCGTGATCATCTGGGCTGGAACAGATCGTGGGATGTACAGCCCGTGAGCATCTG 840
Db      261 ArgTyrIleIleLeuGlnLysLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280
Qy      841 GACATCGCGCAGGCGCCCAAGAGGCCCTTCCGCACTACGTGACCGCTTCTTCAAGACC 900
Db      281 AspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePheLysThr 300
Qy      901 CTGCGCGCGCAGAGGCGCCACCGACGTAAGACTGATGACCGGACCCCTGCTGTG 960
Db      301 LeuArgAlaGlnGlnAlaThrGlnAsnValLysAsnTyrMetThrAspThrLeuVal 320
Qy      961 CAGAACGCCAACCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCGCGCCACCCCTG 1020
Db      321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340
Qy      1021 GAGGAGATGATACCGCTGCGAGGCGCGTGGCGGCGCCCGCCACAAGCGCGTGTG 1080
Db      341 GlnGlnMetMetThrAlaCysGlnGlnValGlyGlyProSerHisLysAlaArgValLeu 360
Qy      1081 GCCGAGGCGATGAGCGCAGGCCACAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140
Db      361 AlaGlnAlaMetSerGlnThrAsnAsnAlaAsnIleMetMetGlnArgSerAsnPheLys 380
Qy      1141 GCGCGCGCGCGCAACGTCAAGTCTTCACTGCGGCAAGAGGGCCACATCGCCCAAGAAC 1200
Db      381 GlyProArgArgIleIleLysCysPheAsnCysGlyLysGlnGlnHisLeuAlaArgAsn 400
Qy      1201 TGC CGCGCGCGCGCGCAAGAGGGCTGTGTAAGTGCAGAGAGGCGCCACAGATGAAG 1260
Db      401 CysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGlnGlnHisGlnMetLys 420
Qy      1261 GACTGCACCGAGCGCGCAGGCCAACTTCTGCGCAAGATGTGCCAGCCACAAGGGCGCG 1320
```

```
Db      421 AspCysThrGlnLysArgGlnAlaAsnPheLeuGlyLysIleTyrProSerHisLysGlyArg 440
Qy      1321 CCCGCAACTTCTGCAAGAACCGCAGCGAGCCCGCGCCCGCCACCGTGCACCGCCCC 1380
Db      441 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454
Qy      1381 CCCGCGAGAGCTTCCGCTTGAAGAGACACCCCGCCCGCCCAAGCAGAGCCCAAGAG 1440
Db      455 ProAlaGlnSerPheArgPheGlnGlnThrThrProAlaProLysGlnGlnProArgGlu 474
Qy      1441 CGCGAGCCCTACCGCGAGCGCCCTGACCGCCTGCGCAGACCTGTTGCGAGCGGCCCTG 1500
Db      475 -----ArgGlnProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490
Qy      1501 AGCCAG 1506
Db      491 SerGln 492
```

```
RESULT 15
ID Q90ME2 PRELIMINARY; PRT; 492 AA.
AC Q90ME2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Gag protein.
GN Name=gag;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21565559; PubMed=11709098; DOI=10.1089/08892220152644241;
RA Engelbrecht S., de Villiers T., Sampson C.C., zur Megede J.,
RA Barnett S.W., van Rensburg B.J.;
RT "Genetic analysis of the complete gag and env genes of HIV type 1
RT subtype C primary isolates from South Africa.";
RL AIDS Res. Hum. Retroviruses 17:1533-1547(2001).
DR EMBL; AF391269; AK77530.1; -.
DR HSSP; P24736; INCP.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zF-CCHC; 2.
DR SMART; SM00343; ZNF_C2HC; 2.
DR PROSITE; PS50158; ZF_CCHC; 2.
KW AIDS; Core protein; Polyprotein.
SQ SEQUENCE 492 AA; 55186 MW; E8023DE89216F43D CRC64;
```

```
Alignment Scores:
Pred. No.: 5.55e-97 Length: 492
Score: 2430.00 Matches: 456
Percent Similarity: 94.82% Conservative: 20
Best Local Similarity: 90.84% Mismatches: 16
Query Match: 85.56% Indels: 10
DB: 2 Gaps: 2
```

US-09-475-704A-4 (1-1509) x Q90ME2 (1-492)

```
Qy      1 ATGGCGCGCGCGCGCAGCATCTGCGCGCGGAGAGCTGACAAGTGGAGAAGATCCGC 60
Db      1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspThrTrpGlnLysIleArg 20
Qy      61 CTGCGCGCGCGCGCGCAAGAGCACTACATGCTGAAGACACTGTGTGGGCCAGCGCGAG 120
Db      21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40
Qy      121 CTGAGGGCTTGCCTTGAACCCCGCGCTGTGTAAGACCGCGAGGGCTGCAAGCATC 180
Db      41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnGlyCysLysGlnIle 60
```

```
QY 181 ATGAAGCAGCTGCAGCCCCCGCTGCAGACCGGCACGAGAGCTGCCGACCTGTACAAC 240
Db 61 ILeaRgInLeuGInProAlaLeuGInThRgLyThRgInGluLeuLySseRLeuTyRAsn 80
QY 241 ACCGTGGCCACCCCTGTACTGCGTGCACGCCGGCATCGAGGTCCCGCACACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyRcysValHisAlaLysIleGluValArgAspThrLysGluAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGACAGACAAGTCCCGACAGAGAAGACCCACAGGCCAAGAG 360
Db 101 LeuAspArgIleGluGluGluGluLysLysCysGlnGlnLysThrGlnGlnAlaLysGlu 120
QY 361 GCCGACGGCAGAGGTGAGCCAGAACTACCCCATCGTGAGAGAACCTGACAGGGCCAGATG 420
Db 121 AlaAspGlyLysIleSerGlnAsnTyRProIleValGlnAsnLeuGlnGlyGlnMetVal 140
QY 421 CACCAAGGCCATCAGCCCCCGCACCCCTGAAAGCGCTGGAGTGATCGAGAGAGAGGCC 480
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 160
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCGCAGGAC 540
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY 541 CTGAACACGATGTTGAACACCGGTGGCGGGCCACAGAGCGCCCATGCGATGCTGAAGGAC 600
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200
QY 601 ACCATCAACGAGAGGGCGCGGAGTGAGGACCGCTGACACCCCGTGACAGCGCGCCGCTG 660
Db 201 ThrIleAsnGluGluAlaAlaGluTrpAspArgIleHisProValHisAlaGlyProIle 220
QY 661 GCGCCCGGCGAGATGCGCGGACCCCGCGGCGAGCGCATCGCCGGCGCCACAGACCTG 720
Db 221 AlaProGlyGlnMetArgGluProArgLysSerAspIleAlaGlyThrThrSerThrLeu 240
QY 721 CAGGAGCAGATCGCTGGATGACCAAGAACCCCGCGTGCCCGCTGGGCGACATTTACAAG 780
Db 241 GlnGluGlnIleAlaTrpMetThrSerAsnProProValProValGlyGluIleTyLys 260
QY 781 CGGTGATCATCTGGGCTGGAACAAGATCGTGCGGATGTACAGCCCGTGAGCATCTG 840
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyRserProValSerIleLeu 280
QY 841 GACATCGCGCAGGGCGCCAGAGCGCTTCCGCGACTACGTGAACCGCTTCTTCAAGACC 900
Db 281 AspIleLysGlnGlyProLysGluProPheArgAspTyValAspArgPhePheLysThr 300
QY 901 CTGCGCGCGAGAGAGCCACCCAGAGCGTGAAGAAGTGAATGATGACCGACCCCTGTG 960
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320
QY 961 CAGAAAGCAACCCGACTGCAAGACCATCTGCGCGCTCTCGCGCCGCGCGCACCTG 1020
Db 321 GlnAsnAlaAsnProAspCysLysIleIleLeuArgGlyLeuGlyProGlyAlaThrLeu 340
QY 1021 GAGGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCGCGCCACAAGCGCGCTG 1080
Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysValaArgValLeu 360
QY 1081 GCCGAGCGGATGAGCCAGGCCAAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140
Db 361 AlaGluAlaMetSerGlnAlaAsnSerThrAsnIleMetMetGlnArgGlyAsnPhelLys 380
QY 1141 GCGCCCGCGGCAAGTCAAGTGTCTTCAACTGCGGCAAGAGGCGCCATGCGCAAGAAC 1200
Db 381 GlyProLysArgAsnIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArgAsn 400
QY 1201 TGCGCGCGCGCGCAAGAGGCTGCTGAAGTGGCGCAAGAGGCGCCACAGATGAAG 1260
Db 401 TyrArgAlaProArgLysLysValCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420
```

```
QY 1261 GACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGAGCCAGCCACAAGGCGCC 1320
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGGCAACTTCTGCGAGAACCGGAGCGAGCCCGCGCCCGCCACCGGTGCCACCGCCGCC 1380
Db 441 ProGlyAsnPheLeuGlnAsnArgThrGlu-----ProThrAlaPro 454
QY 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCCCGCCCAAGAGAGCCCAAGAC 1440
Db 455 ProAlaGluSerPheArgPheGluGluThrAsnProAlaProLysGlnGluProLysAsp 474
QY 1441 CGCGAGCCCTACCGCGAGCGCCCTGACCGCCCTGGCGAGCCTGTTCGGCAGCGCGCCCTG 1500
Db 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProSer 490
QY 1501 AGCCAG 1506
Db 491 SerGln 492
```

Search completed: March 11, 2005, 15:39:49  
Job time : 289.58 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 11:36:50 ; Search time 857.019 Seconds  
(without alignments)  
10423.210 Million cell updates/sec

Title: US-09-475-704A-4  
Perfect score: 1509  
Sequence: 1 atggcgccgcgcgcgcagcat.....gcggcccccctgagccagtaa 1509

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1509	100.0	1509	3	AAAS1610 HIV synth
2	1509	100.0	1509	6	AA144549 HIV-1 p55
3	1509	100.0	1509	6	ABL39954 Synthetic
4	1509	100.0	1509	12	ADM73759 HIV-1 pol
5	1491.4	98.8	1509	6	ABL39958 Synthetic
6	1491.4	98.8	1509	12	ADM73763 HIV-1 pol
7	1489.8	98.7	1509	3	AAAS1626 HIV codon
8	1489.8	98.7	1509	6	AA144553 HIV-1 p55
9	1297.6	86.0	1494	6	ABL39972 Synthetic
10	1297.6	86.0	1494	12	ADM73785 HIV-1 pol
11	1294.8	85.8	1491	6	ABL40020 Synthetic
12	1294.8	85.8	1491	12	ADM73833 HIV-1 pol
13	1276.8	84.6	1479	3	AAAS1609 HIV synth
14	1276.8	84.6	1479	6	AA144548 HIV-1 p55
15	1276.8	84.6	1479	6	ABL39953 Synthetic
16	1276.8	84.6	1479	12	ADM73758 HIV-1 pol
17	1275.2	84.5	3162	10	ADCI3239 DNA of HI
18	1275.2	84.5	3462	10	ADCI3237 DNA of HI
19	1275.2	84.5	4419	8	ACA03523 Synthetic
20	1275.2	84.5	4419	10	ADCI3240 DNA of HI

21	1275.2	84.5	4615	10	ADCI3257 DNA of HI
22	1275.2	84.5	4702	10	ADCI3259 DNA of HI
23	1275.2	84.5	4716	8	ACA03522 Synthetic
24	1275.2	84.5	4716	10	ADCI3238 DNA of HI
25	1273.6	84.4	2742	8	ACA03524 Synthetic
26	1273.6	84.4	2742	10	ADCI3241 DNA of HI
27	1273.6	84.4	3930	10	ADCI3230 DNA of HI
28	1273.6	84.4	3930	10	ADCI3231 DNA of HI
29	1273.6	84.4	3930	10	ADCI3232 DNA of HI
30	1273.6	84.4	5145	8	ACA03521 Synthetic
31	1273.6	84.4	5145	10	ADCI3233 DNA of HI
32	1272.6	84.3	4713	8	ACA03592 Synthetic
33	1272.6	84.3	4713	10	ADCI3280 DNA of HI
34	1271.6	84.3	5184	8	ACA03591 Synthetic
35	1271.6	84.3	5184	10	ADCI3279 DNA of HI
36	1270.2	84.2	2742	8	ACA03590 Synthetic
37	1270.2	84.2	2742	10	ADCI3278 DNA of HI
38	1260.8	83.6	1479	3	AAAS1625 HIV codon
39	1260.8	83.6	1479	6	ABL39957 Synthetic
40	1260.8	83.6	1479	12	ADM73762 HIV-1 pol
41	1259.2	83.4	1479	6	AA144552 HIV-1 p55
42	1235.4	81.9	1545	12	ADN36395 Human gen
43	1234.4	81.8	4546	10	ADCI3255 DNA of HI
44	1232.2	81.7	8186	12	ADN36401 Human gen
45	1202.2	79.7	4288	4	AAH20868 HIV-1 sub

## ALIGNMENTS

RESULT 1  
ID AAAS1610 standard; DNA; 1509 BP.  
XX  
AC AAAS1610;  
XX  
DT 31-OCT-2000 (first entry)  
XX  
DE HIV synthetic Gag polynucleotide.  
XX  
KW Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;  
KW DNA immunization; packaging cell line; antigen presentation; ss.  
XX  
OS Human immunodeficiency virus; type C strain AF110967.  
OS Synthetic.  
XX  
PN WO200039304-A2.  
XX  
PD 06-JUL-2000.  
XX  
PF 30-DEC-1999; 99WO-US031273.  
XX  
PR 31-DEC-1998; 98US-0114495P.  
PR 01-SEP-1999; 99US-0152195P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Barnett S, Zur Megede J;  
PI WPI; 2000-452401/39.  
XX  
DR  
XX  
PT Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env  
PT polypeptide and the polypeptide useful for immunizing a mammal especially  
PT human against HIV.  
XX  
PS Claim 2; Page 93; 113pp; English.  
XX  
CC Expression cassettes comprising a polynucleotide encoding antigenic type  
CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
CC in DNA immunization, generation of packaging cell lines and production of  
CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
CC cassettes exhibit increased potency for induction of cytotoxic T-  
CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-

CC assemble into non-infectious virus-like particles which are used as a  
CC matrix for the proper presentation of an antigen entrapped or associated  
CC to the immune system of the host

Sequence 1509 BP; 320 A; 556 C; 472 G; 161 T; 0 U; 0 Other;

Query Match	100.0%;	Score 1509;	DB 3;	Length 1509;
Best Local Similarity	100.0%;	Pred. No. 5.3e-177;		
Matches 1509; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	ATGGGGCGCCCCCGCCAGCATCTGCGCGGCGGAGAACTGGACAAGTGGGAGAAAGATCCGC	60
Db	1	ATGGGCGCCCCCGCCAGCATCTGCGCGGCGGAGAACTGGACAAGTGGGAGAAAGATCCGC	60
QY	61	CTGCGCCCCCGGCGGCAAGAACCTACATGCTGAAGCACCTGCTGGGCCAGCCGAG	120
Db	61	CTGCGCCCCCGGCGGCAAGAACCTACATGCTGAAGCACCTGCTGGGCCAGCCGAG	120
QY	121	CTGGAGGGCTTCGCCCTGTAACCCCGGCTGCTGGAGACCCGCGAGGGCTGCAAGCATC	180
Db	121	CTGGAGGGCTTCGCCCTGTAACCCCGGCTGCTGGAGACCCGCGAGGGCTGCAAGCATC	180
QY	181	ATGAAGCAGCTGCAGCCCGCCCTGCAGACCGGCACCGAGAGCTGCGCAGCCTGTACAA	240
Db	181	ATGAAGCAGCTGCAGCCCGCCCTGCAGACCGGCACCGAGAGCTGCGCAGCCTGTACAA	240
QY	241	ACCGTGGCCACCCCTGTACTGCGGTGCACGCGCGCATCGAGGTCGCGACACCAAGAGGCC	300
Db	241	ACCGTGGCCACCCCTGTACTGCGGTGCACGCGCGCATCGAGGTCGCGACACCAAGAGGCC	300
QY	301	CTGCACAGATCGAGAGGAGCAGAACAGTCCCAAGACAAGACCCAGCAGGCCAAGAG	360
Db	301	CTGCACAGATCGAGAGGAGCAGAACAGTCCCAAGACAAGACCCAGCAGGCCAAGAG	360
QY	361	GCCGACGGCAAGGTGAGCCAGACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGCTG	420
Db	361	GCCGACGGCAAGGTGAGCCAGACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGCTG	420
QY	421	CACCAAGCCATCAGCCCCCGCACCCCTGAACGCTGGTGAAGGTGATCGAGAGAGAGGCC	480
Db	421	CACCAAGCCATCAGCCCCCGCACCCCTGAACGCTGGTGAAGGTGATCGAGAGAGAGGCC	480
QY	481	TTCAGCCCCGAGGTGATCCCATGTTACCGCCCTGAGCGAGGGCGCCACCCCCCAGAG	540
Db	481	TTCAGCCCCGAGGTGATCCCATGTTACCGCCCTGAGCGAGGGCGCCACCCCCCAGAG	540
QY	541	CTGAACAGATGTTGAACACCGTGGCGGCCACAGGCCGCATGACAGATGCTGAAGGAC	600
Db	541	CTGAACAGATGTTGAACACCGTGGCGGCCACAGGCCGCATGACAGATGCTGAAGGAC	600
QY	601	ACCATCAACGAGAGGCGCGGAGTGGACCGCCTGCACCCCGTGACGGCGGCCCGTG	660
Db	601	ACCATCAACGAGAGGCGCGGAGTGGACCGCCTGCACCCCGTGACGGCGGCCCGCGTG	660
QY	661	GCCCCCGGCGAGATGCGCGACCCCGCGGCAAGCAATCGCCGCGGCCACAGCACCTG	720
Db	661	GCCCCCGGCGAGATGCGCGACCCCGCGGCAAGCAATCGCCGCGGCCACAGCACCTG	720
QY	721	CAGAGCAGATCGCCTGATGACCAAGCAACCCCGCTGCCCCGTGGCGCATCTACAAG	780
Db	721	CAGAGCAGATCGCCTGATGACCAAGCAACCCCGCTGCCCCGTGGCGCATCTACAAG	780
QY	781	CGGTGGATCATCTGGGGCTGAACAAGATCGTGCGATGTACAGCCCGTGAGCATCTG	840
Db	781	CGGTGGATCATCTGGGGCTGAACAAGATCGTGCGATGTACAGCCCGTGAGCATCTG	840
QY	841	GACATCCGCCAAGGCGCCCAAGAGCCCTTCGCGCACTACGTGAACGCGCTTCTTAAGAC	900
Db	841	GACATCCGCCAAGGCGCCCAAGAGCCCTTCGCGCACTACGTGAACGCGCTTCTTAAGAC	900
QY	901	CTGCGCGCCGAGCAGGCCACCCAGACGTGAAGAACTGATGACCGAGACCCCTGCTGGT	960
Db	901	CTGCGCGCCGAGCAGGCCACCCAGACGTGAAGAACTGATGACCGAGACCCCTGCTGGT	960

QY	961	CAGAACGCCAACCCCGCACTGCAAGACCATCCTCGCGCTCTCGGCCCCGGCGCAACCTTG	1020
Db	961	CAGAACGCCAACCCCGCACTGCAAGACCATCCTCGCGCTCTCGGCCCCGGCGCAACCTTG	1020
QY	1021	GAGGAGATGATGACCGCCTTGCCAGGGCGTGGGGGGCCCCGGCCACAAGGCCCGGTGCTG	1080
Db	1021	GAGGAGATGATGACCGCCTTGCCAGGGCGTGGGGGGCCCCGGCCACAAGGCCCGGTGCTG	1080
QY	1081	GCCGAGCGGATGAGGCCAGGCCCAACGCGTGAACATCATGATGCAAGAGCAACTTCAAG	1140
Db	1081	GCCGAGCGGATGAGGCCAGGCCCAACGCGTGAACATCATGATGCAAGAGCAACTTCAAG	1140
QY	1141	GGCCCCGGCGGCAACGTCAAGTGTCTCACTGCGGCAAGAGAGGGCCACATGCGCCAAGAAC	1200
Db	1141	GGCCCCGGCGGCAACGTCAAGTGTCTCACTGCGGCAAGAGAGGGCCACATGCGCCAAGAAC	1200
QY	1201	TGCGCGCGCCCCCGCCCAAGAAAGGCGTGTGAAAGTGGCGCAAGAGGGCCACCATGAAG	1260
Db	1201	TGCGCGCGCCCCCGCCCAAGAAAGGCGTGTGAAAGTGGCGCAAGAGGGCCACCATGAAG	1260
QY	1261	GACTGCACCGGAGCGCCAGGCCCACTTCTGGGCAAGATCTGGCCCAAGGCCCAAGGGCGCG	1320
Db	1261	GACTGCACCGGAGCGCCAGGCCCACTTCTGGGCAAGATCTGGCCCAAGGCCCAAGGGCGCG	1320
QY	1321	CCGGGCAACTTCTGCAAGAACCGCAGCAGCGCGCGCCCCCAACCGTGCACCGCCCCC	1380
Db	1321	CCGGGCAACTTCTGCAAGAACCGCAGCAGCGCGCGCCCCCAACCGTGCACCGCCCCC	1380
QY	1381	CCCGCCGAGAGCTTCCGTTGAGAGAGACCACCCCCGCCCCCAAGCAGAGGCCCAAGAAC	1440
Db	1381	CCCGCCGAGAGCTTCCGTTGAGAGAGACCACCCCCGCCCCCAAGCAGAGGCCCAAGAAC	1440
QY	1441	CGGAGCCCTACCGCGAGCCCTGACCGCCTGCGCAGCCTGTTGGCAGCGGCCCTCTG	1500
Db	1441	CGGAGCCCTACCGCGAGCCCTGACCGCCTGCGCAGCCTGTTGGCAGCGGCCCTCTG	1500
QY	1501	AGCCAGTAA 1509	
Db	1501	AGCCAGTAA 1509	

Accession	Source	Organism	Accession	Source	Organism
AL44549	standard; DNA; 1509 BP.		AL44549	standard; DNA; 1509 BP.	
AC	AA44549;		AC	AA44549;	
DT	29-AUG-2003 (revised)		DT	29-AUG-2003 (revised)	
DT	08-NOV-2002 (first entry)		DT	08-NOV-2002 (first entry)	
DE	HIV-1 p55gag polypeptide coding sequence 2.		DE	HIV-1 p55gag polypeptide coding sequence 2.	
KM	HIV; ds; vaccine; gene; immune response; microparticle;		KM	HIV; ds; vaccine; gene; immune response; microparticle;	
KM	adsorbent surface; poly(alpha-hydroxy acid); polyhydroxy butyric acid;		KM	adsorbent surface; poly(alpha-hydroxy acid); polyhydroxy butyric acid;	
KM	polycaprolactone; polyorthoester; polycyanacrylate; detergent;		KM	polycaprolactone; polyorthoester; polycyanacrylate; detergent;	
KM	submicron emulsion; viral infection; bacterial infection;		KM	submicron emulsion; viral infection; bacterial infection;	
KM	parasitic infection; HIV-1 p55gag polypeptide.		KM	parasitic infection; HIV-1 p55gag polypeptide.	
OS	Human immunodeficiency virus 1.		OS	Human immunodeficiency virus 1.	
PN	WO200226209-A2.		PN	WO200226209-A2.	
PD	04-APR-2002.		PD	04-APR-2002.	
PF	28-SEP-2001; 2001WO-US030540.		PF	28-SEP-2001; 2001WO-US030540.	
PR	28-SEP-2000; 2000US-0236105P.		PR	28-SEP-2000; 2000US-0236105P.	
PR	30-AUG-2001; 2001US-0315905P.		PR	30-AUG-2001; 2001US-0315905P.	
PA	(CHIR ) CHIRON CORP.		PA	(CHIR ) CHIRON CORP.	
OT	O'hagan D, Otten G, Donnelly JT, Polo JM, Barnett S, Singh M;		OT	O'hagan D, Otten G, Donnelly JT, Polo JM, Barnett S, Singh M;	

PI Ulmer J, Dubensky TW;  
XX  
DR WPI, 2002-519084/55.  
XX  
PT A microparticle to which a biologically active macromolecule is adsorbed,  
PT for use as a vaccine composition to treat viral, bacterial or parasitic  
PT infections, comprises a polymer microparticle, a detergent and a  
PT submicron emulsion.  
XX  
PS Claim 72; Fig 2; 100pp; English.  
XX  
CC The invention relates to a method of raising an immune response in a host  
CC animal. The method of the invention comprises administering a  
CC microparticle that has an adsorbent surface to which a first biologically  
CC active macromolecule (e.g. a nucleic acid) has been adsorbed. The  
CC microparticle comprises a polymer microparticle of poly(alpha-hydroxy  
CC acid), a polyhydroxy butyric acid, a polycaprolactone, a polyorthoester,  
CC a polycyanosacrylate, a detergent, and submicron emulsion. The method/  
CC microparticle of the invention is useful for immunising a host animal  
CC against viral, bacterial or parasitic infections. The present DNA  
CC sequence encodes a HIV-1 p55gag polypeptide. (Updated on 29-AUG-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 1509 BP; 320 A; 556 C; 472 G; 161 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1509; DB 6; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 5.3e-177;  
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGCGCCGCGCCAGCATCTGCGCGCGGAGAGCTGGACAAGTGGAGAAAGATCCGC 60  
DB 1 ATGGGCGCCGCGCCAGCATCTGCGCGCGGAGAGCTGGACAAGTGGAGAAAGATCCGC 60  
QY 61 CTGCGCCCCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGTGGCCAGCCGCGAG 120  
DB 61 CTGCGCCCCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGTGGCCAGCCGCGAG 120  
QY 121 CTGGAAGGGCTTCGCGCTGAAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCAGATC 180  
DB 121 CTGGAAGGGCTTCGCGCTGAAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGCAGCCCCCGCTGCAAGCCGCGCAGAGAGCTGGCGACGCTGTACAAC 240  
DB 181 ATGAAGCAGCTGCAGCCCCCGCTGCAAGCCGCGCAGAGAGCTGGCGACGCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGACGCGCGCATGAGGTCCGCAACCAAGAGGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGCGTGACGCGCGCATGAGGTCCGCAACCAAGAGGGCC 300  
QY 301 CTGGAACAAGATCGAGAGGAGCAGAACAAGTCCCAAGAGAAGACCAGCAGGCCAAGAG 360  
DB 301 CTGGAACAAGATCGAGAGGAGCAGAACAAGTCCCAAGAGAAGACCAGCAGGCCAAGAG 360  
QY 361 GCGGACGGCAAGGTGAGCCAGAACTAACCCCATCTGTGAGAACTTGAGGGCCAGATG 420  
DB 361 GCGGACGGCAAGGTGAGCCAGAACTAACCCCATCTGTGAGAACTTGAGGGCCAGATG 420  
QY 421 CACCAAGCCATCAGCCCCCGGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
DB 421 CACCAAGCCATCAGCCCCCGGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
QY 481 TTCAGCCCCGAGGTGATCCCATATTTACCGGCTGTAGCGAGGGCCACCCCGCAGAG 540  
DB 481 TTCAGCCCCGAGGTGATCCCATATTTACCGGCTGTAGCGAGGGCCACCCCGCAGAG 540  
QY 541 CTGAACAACGATGTTGAACACCGTGGCGCGCCACCAAGCCCATGAGATGCTGAAGAG 600  
DB 541 CTGAACAACGATGTTGAACACCGTGGCGCGCCACCAAGCCCATGAGATGCTGAAGAG 600  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGGAGCCGCTGTGACCCCGTGTGAGCGCGCCCG 660  
DB 601 ACCATCAACGAGAGGCGCGCGAGTGGGAGCCGCTGTGACCCCGTGTGAGCGCGCCCG 660

QY 661 GCGCCCGGCGCAGATGCGCGGACCCCGCGGACGCAATCGCCGCGCCACCAAGACCTG 720  
DB 661 GCGCCCGGCGCAGATGCGCGGACCCCGCGGACGCAATCGCCGCGCCACCAAGACCTG 720  
QY 721 CAGGACGAGATCGCTGTGATGACCAACAACCCCGGTGCGCGGTGGGGGACATCTAACAG 780  
DB 721 CAGGACGAGATCGCTGTGATGACCAACAACCCCGGTGCGCGGTGGGGGACATCTAACAG 780  
QY 781 CCGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGCTGAGCATCTG 840  
DB 781 CCGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGCTGAGCATCTG 840  
QY 841 GACATCCGCGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTAAGACC 900  
DB 841 GACATCCGCGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTAAGACC 900  
QY 901 CTGCGCGCGGAGCAGGCGCAACCGAAGAGCTGAAGAACTGATGACCGAGACCTGTG 960  
DB 901 CTGCGCGCGGAGCAGGCGCAACCGAAGAGCTGAAGAACTGATGACCGAGACCTGTG 960  
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTTCCGCGCGCGCGCACCTG 1020  
DB 961 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTTCCGCGCGCGCGCACCTG 1020  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGGCGGCGCCGCGCAAGGCGCGCTGCTG 1080  
DB 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGGCGGCGCCGCGCAAGGCGCGCTGCTG 1080  
QY 1081 GCGGAGCGGATGAGCCAGGCGCAACGCTGAACATCATGATGAGAAGCAACTTCAAG 1140  
DB 1081 GCGGAGCGGATGAGCCAGGCGCAACGCTGAACATCATGATGAGAAGCAACTTCAAG 1140  
QY 1141 GCGCCCGGCGCAAGCTCAAGTCTTCAACTGCGGCAAGAGGCGCAATCGCCAGAAG 1200  
DB 1141 GCGCCCGGCGCAAGCTCAAGTCTTCAACTGCGGCAAGAGGCGCAATCGCCAGAAG 1200  
QY 1201 TGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGGCGCCACCAAGTGAAG 1260  
DB 1201 TGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGGCGCCACCAAGTGAAG 1260  
QY 1261 GACTGCACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCTGGCCAGCAAGAGGCGCG 1320  
DB 1261 GACTGCACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCTGGCCAGCAAGAGGCGCG 1320  
QY 1321 CCGGCAACTTCTGCAAGACCGCAGAGCGCGCGCGCCCGCCACCGTGCCACGCGCCCC 1380  
DB 1321 CCGGCAACTTCTGCAAGACCGCAGAGCGCGCGCGCGCCCGCCACCGTGCCACGCGCCCC 1380  
QY 1381 CCGCGCAGAGCTTCGCTTGAAGAGACCAACCCCGCGCCCAAGAGGCGCAAGAGAC 1440  
DB 1381 CCGCGCAGAGCTTCGCTTGAAGAGACCAACCCCGCGCCCAAGAGGCGCAAGAGAC 1440  
QY 1441 CCGGAGCCCTACCGGAGGCGCCCTGACCGGCTGGCGCAGCTGTTGGCAGCGGCGCCCTG 1500  
DB 1441 CCGGAGCCCTACCGGAGGCGCCCTGACCGGCTGGCGCAGCTGTTGGCAGCGGCGCCCTG 1500  
QY 1501 AGCCAGTAA 1509  
DB 1501 AGCCAGTAA 1509

RESULT 3  
ABL39954  
ID ABL39954 standard; DNA; 1509 BP.  
XX ABL39954;  
AC  
XX  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DB Synthetic Gag polynucleotide sequence SEQ ID NO:4.  
XX  
XX Human immunodeficiency virus type C; antigenic HIV type C protein;  
KW immunogenic; gag; pol; vif; vpr; tat; rev; env; nef;

KM immunostimulant; gene therapy; gene; ds.  
XX  
OS Human immunodeficiency virus; type C.  
OS Synthetic.  
XX  
PN WO200204493-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megele J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
XX  
DR WPI; 2002-154920/20.  
XX  
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunization or generation of packaging cell  
PT lines, particularly in gene therapy.  
XX  
PS Example 1; Fig 2; 233bp; English.  
XX  
CC The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (1). (1) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC Type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 1509 BP; 320 A; 556 C; 472 G; 161 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1509; DB 6; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 5.3e-177;  
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGCGCCCGCGCCAGCATCTCGCGCGGAGAGAGCTGGACAAGTGGAGAGATCCGC 60  
Db 1 ATGGGCGCCCGCGCCAGCATCTCGCGCGGAGAGAGCTGGACAAGTGGAGAGATCCGC 60  
QY 61 CTGCGCCCCCGCGGCAAGAGCACTAATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
Db 61 CTGCGCCCCCGCGGCAAGAGCACTAATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
QY 121 CTGGAAGGGCTTCGCGCTGAACCCCGGCTGCTGAGACCGCGAGGGCTGCAAGCATC 180  
Db 121 CTGGAAGGGCTTCGCGCTGAACCCCGGCTGCTGAGACCGCGAGGGCTGCAAGCATC 180  
QY 181 ATGAAGCAGCTGAGCCCGGCTGAGACCGGCAACGAGAGCTGCGAGCCTGTACAAC 240  
Db 181 ATGAAGCAGCTGAGCCCGGCTGAGACCGGCAACGAGAGCTGCGAGCCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCTGACGCGCGCATCGAGTCCGCGACCAAGAGAGCC 300  
Db 241 ACCGTGGCCACCTGTACTGCTGACGCGCGCATCGAGTCCGCGACCAAGAGAGCC 300  
QY 301 CTGGAACAAGATCGAGAGAGAGAGACAAGTCCAGAGAGAGAGAGAGAGAGAGAG 360  
Db 301 CTGGAACAAGATCGAGAGAGAGAGACAAGTCCAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTTGAGAGGCCAGATG 420  
Db 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTTGAGAGGCCAGATG 420  
QY 421 CACCAAGGCATACGCCCCCGGACCCCTGAAGCGCTGGTGAAGGTGATCGAGAGAGGCC 480

Db 421 CACCAAGGCATACGCCCCCGGACCCCTGAAGCGCTGGTGAAGGTGATCGAGAGAGGCC 480  
QY 481 TTACAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGAGGAGCGCCACCCCCAGAG 540  
Db 481 TTACAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGAGGAGCGCCACCCCCAGAG 540  
QY 541 CTGAACAAGATGTTGAACAACCGTGGGCGGCCACAGGCGCCCATGAGATGCTGAAGAG 600  
Db 541 CTGAACAAGATGTTGAACAACCGTGGGCGGCCACAGGCGCCCATGAGATGCTGAAGAG 600  
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGACCGGCTGCAACCCGCTGAGAGCGCGCCG 660  
Db 601 ACCATCAACGAGAGAGCGCGCGAGTGGACCGGCTGCAACCCGCTGAGAGCGCGCCG 660  
QY 661 GCCCGCGGCGAGATGCGCGACCCCGCGGAGAGCATCGCGCGGCCACCAAGACCCCTG 720  
Db 661 GCCCGCGGCGAGATGCGCGACCCCGCGGAGAGCATCGCGCGGCCACCAAGACCCCTG 720  
QY 721 CAGGAGCAGATGCGCTGATGACCAAGCAACCCCGCTGCGCGTGGGCGCATCTACAAG 780  
Db 721 CAGGAGCAGATGCGCTGATGACCAAGCAACCCCGCTGCGCGTGGGCGCATCTACAAG 780  
QY 781 CGGTGATCATCTGGGCGCTGAACAAGATGCTGCGGATGTAAGCCCGTGAAGCATCTG 840  
Db 781 CGGTGATCATCTGGGCGCTGAACAAGATGCTGCGGATGTAAGCCCGTGAAGCATCTG 840  
QY 841 GACATCCGCGGAGGCGCCCAAGAGAGCCCTTCGCGAGTACGTGAGACCGCTTCTCAAG 900  
Db 841 GACATCCGCGGAGGCGCCCAAGAGAGCCCTTCGCGAGTACGTGAGACCGCTTCTCAAG 900  
QY 901 CTGCGCGCGCGAGGCGCCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 901 CTGCGCGCGCGAGGCGCCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 CAGAACGCGCAACCCCGAGCTGCAAGACCATCTGCGCGCTCTGCGCGCGCGCGCAC 1020  
Db 961 CAGAACGCGCAACCCCGAGCTGCAAGACCATCTGCGCGCTCTGCGCGCGCGCGCAC 1020  
QY 1021 GAGGAGATGATGACCGCGCTGCGAGGCGGTGGGCGCGCGCGCGCGCGCGCGCTG 1080  
Db 1021 GAGGAGATGATGACCGCGCTGCGAGGCGGTGGGCGCGCGCGCGCGCGCGCGCTG 1080  
QY 1081 GCCGAGCGGATGAGCGCCAGGCGCAACAGCGTGAACATGATGACAGAGAGCACTTCA 1140  
Db 1081 GCCGAGCGGATGAGCGCCAGGCGCAACAGCGTGAACATGATGACAGAGAGCACTTCA 1140  
QY 1141 GGCCCCCGGCGCAACGTCATAGTGTCTCACTGCGGCAAGAGGCGCCATGCGCAAGAA 1200  
Db 1141 GGCCCCCGGCGCAACGTCATAGTGTCTCACTGCGGCAAGAGGCGCCATGCGCAAGAA 1200  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGCCACAGATGAAG 1260  
Db 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGCCACAGATGAAG 1260  
QY 1261 GACTGACCGGAGCGCGCGCAACTTCTGCGCAAGATCTGGCGCAAGAGGCGCGC 1320  
Db 1261 GACTGACCGGAGCGCGCGCAACTTCTGCGCAAGATCTGGCGCAAGAGGCGCGC 1320  
QY 1321 CCGGCAACTTCTGCGCAACCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
Db 1321 CCGGCAACTTCTGCGCAACCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
QY 1381 CCGGCGAGAGCTTCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGCG 1440  
Db 1381 CCGGCGAGAGCTTCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGCG 1440  
QY 1441 CGCGAGCCCTACCGCGAGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
Db 1441 CGCGAGCCCTACCGCGAGCGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
QY 1501 AGCCAGTAA 1509  
|||||||

Db 1501 AGCCAGTAA 1509

RESULT 4  
ADM73759  
ID ADM73759 standard; DNA; 1509 BP.  
AC ADM73759;  
XX  
XX 03-JUN-2004 (first entry)  
DT  
XX  
DE HIV-1 polynucleotide #2.  
XX  
KM HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
KW HIV type C protein; immunostimulant.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN US2003223961-A1.  
XX  
PD 04-DEC-2003.  
XX  
PF 05-JUL-2001; 2001US-00899575.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (MEGE/) MEGEDE J Z.  
PA (BARN/) BARNETT S W.  
PA (ENGE/) ENGELBRECHT S.  
PA (RENS/) RENSBURG E J V.  
XX  
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;  
XX  
DR WPI; 2004-060515/06.  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding an  
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
PT immunization, generating of packaging cell lines or in producing HIV Type  
PT C proteins.  
XX  
PS Example 1; SEQ ID NO 4; 160pp; English.  
XX  
CC The invention relates to an expression cassette comprising a  
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
CC also relates to a recombinant expression system for use in a host cell  
CC comprising an expression cassette, where the polynucleotide sequence  
CC further comprises control elements capable of driving expression in the  
CC selected host cell, a cell comprising an expression cassette where the  
CC polynucleotide sequence further comprises control elements compatible  
CC with the expression in the cell and a composition for generating an  
CC immunological response, comprising an expression cassette. The expression  
CC cassette and the methods of the invention are useful in eliciting an  
CC immune response, in DNA immunisation, in generation of packaging cell  
CC lines and in producing HIV Type C proteins. This sequence represents an  
CC HIV-1 polynucleotide of the invention.  
XX  
SQ Sequence 1509 BP; 320 A; 556 C; 472 G; 161 T; 0 U; 0 Other;

Query Match 100.0%; Score 1509; DB 12; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 5.3e-177;  
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCAGCATCTCGCGCGCGAGAGCTGGACAAGTGGAGAAGATCCGC 60  
DB 1 ATGGGCGCGCGCGCAGCATCTCGCGCGCGAGAGCTGGACAAGTGGAGAAGATCCGC 60  
QY 61 CTGGCGCGCGCGCGCAAGACATCTGTAAGCACTGTGTGGGCCAGCCCGAG 120  
DB 61 CTGGCGCGCGCGCGCAAGACATCTGTAAGCACTGTGTGGGCCAGCCCGAG 120  
QY 121 CTGAGAGGGCTTCGCGCTGTAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCATC 180  
DB 121 CTGAGAGGGCTTCGCGCTGTAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCATC 180

QY 181 ATGAAGCAGCTGCAGCCCGCGCTGCAGACCGGCAACGAGAGTGCAGCCTGTACAAC 240  
DB 181 ATGAAGCAGCTGCAGCCCGCGCTGCAGACCGGCAACGAGAGTGCAGCCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGCGCGCATTCAGGTCCGCGACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGCGTGCACGCGCGCATTCAGGTCCGCGACCAAGAGGCC 300  
QY 301 CTGACAAGATCGAGAGAGGAGCAACAAGTCCCGCAGAAAGCCAGAGCCAGAGAG 360  
DB 301 CTGACAAGATCGAGAGAGGAGCAACAAGTCCCGCAGAAAGCCAGAGCCAGAGAG 360  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGAGAACTGCAGGGCCAGATGTTG 420  
DB 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGAGAACTGCAGGGCCAGATGTTG 420  
QY 421 CACGAGGCCATCAGCCCCCGCACCTGAAACGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
DB 421 CACGAGGCCATCAGCCCCCGCACCTGAAACGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
QY 481 TTCAGCCCCGAGGTGATCCCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCGAGAC 540  
DB 481 TTCAGCCCCGAGGTGATCCCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCGAGAC 540  
QY 541 CTGAACACGATGTTGAACACCCGTGGCGGCCACCAAGCCCATGCAATGCTGAAGAGAC 600  
DB 541 CTGAACACGATGTTGAACACCCGTGGCGGCCACCAAGCCCATGCAATGCTGAAGAGAC 600  
QY 601 ACCATCAACGAGAGGGCGCGCGAGTGGGACCGCTGCACCCCGCTGACCGCCGCGCGCTG 660  
DB 601 ACCATCAACGAGAGGGCGCGCGAGTGGGACCGCTGCACCCCGCTGACCGCCGCGCGCTG 660  
QY 661 GCCCGCGCGCAAGTGGCGGACCCCGCGCGGAGCGACATCGCCGCGCGCCACCAAGCCTG 720  
DB 661 GCCCGCGCGCAAGTGGCGGACCCCGCGCGGAGCGACATCGCCGCGCGCCACCAAGCCTG 720  
QY 721 CAGAGCAGATCGCTGTGATGACCAAGCAACCCCGCTGCGCGCGGAGCATCTACAAG 780  
DB 721 CAGAGCAGATCGCTGTGATGACCAAGCAACCCCGCTGCGCGCGGAGCATCTACAAG 780  
QY 781 CGGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTG 840  
DB 781 CGGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTG 840  
QY 841 GACATCCGCGAGGCGCCCAAGAGCCCTTCGCGCACTACGTGACCGCTTCTCAAGACC 900  
DB 841 GACATCCGCGAGGCGCCCAAGAGCCCTTCGCGCACTACGTGACCGCTTCTCAAGACC 900  
QY 901 CTGGCGCGCGAGCGCCACCCAGGACGCTGAAAGAACTGATGACCGAGACCCTGCTGTG 960  
DB 901 CTGGCGCGCGAGCGCCACCCAGGACGCTGAAAGAACTGATGACCGAGACCCTGCTGTG 960  
QY 961 CAGAACGCCAACCCGACTGCAAGACCATCTGCGCGCTTCTGCGCCCGCGCGCACCTG 1020  
DB 961 CAGAACGCCAACCCGACTGCAAGACCATCTGCGCGCTTCTGCGCCCGCGCGCACCTG 1020  
QY 1021 GAGGAGATGATGACCGCTGCGAGGCGTGGGCGCGCCCGCGCAAGGCCGCGTGTG 1080  
DB 1021 GAGGAGATGATGACCGCTGCGAGGCGTGGGCGCGCCCGCGCAAGGCCGCGTGTG 1080  
QY 1081 GCCGAGGCGATGAGCCAGGCGCAACGCTGAACATCATGATGAGAAAGCAACTTCAAG 1140  
DB 1081 GCCGAGGCGATGAGCCAGGCGCAACGCTGAACATCATGATGAGAAAGCAACTTCAAG 1140  
QY 1141 GCGCGCGCGCGCAAGCTCAAGTGTCTCAACTGCGCGCAAGGAGGCCACATGCGCAAGAAC 1200  
DB 1141 GCGCGCGCGCGCGCAAGCTCAAGTGTCTCAACTGCGCGCAAGGAGGCCACATGCGCAAGAAC 1200  
QY 1201 TGCGCGCGCGCGCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGGGCCACCATGTAAG 1260  
DB 1201 TGCGCGCGCGCGCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGGGCCACCATGTAAG 1260

QY 1261 GACTGCACTGAGCGCCGCAACTTCTTGCGCAAGATCTGGCCCAAGGCGCCG 1320  
Db 1261 GACTGCACTGAGCGCCGCAACTTCTTGCGCAAGATCTGGCCCAAGGCGCCG 1320  
QY 1321 CCGCGCAACTTCTTGCGCAAGCGCAGGAGCCCGCGCCCGCCACCGTGCCACCGCGCCG 1380  
Db 1321 CCGCGCAACTTCTTGCGCAAGCGCAGGAGCCCGCGCCCGCCACCGTGCCACCGCGCCG 1380  
QY 1381 CCGCGCGAGAGCTTCCGCTTGAGGAGACCAACCCCGCGCCCGCAAGCAGAGCCCAAGGAC 1440  
Db 1381 CCGCGCGAGAGCTTCCGCTTGAGGAGACCAACCCCGCGCCCGCAAGCAGAGCCCAAGGAC 1440  
QY 1441 CGCGAGCCCTACCGCGAGCCCTTGACCGCCCTTGCGCAGCCTGTTCCGCAAGCGCCCTG 1500  
Db 1441 CGCGAGCCCTACCGCGAGCCCTTGACCGCCCTTGCGCAGCCTGTTCCGCAAGCGCCCTG 1500  
QY 1501 AGCCAGTAA 1509  
Db 1501 AGCCAGTAA 1509

RESULT 5  
ABL39958  
ID ABL39958 standard; DNA; 1509 BP.

XX ABL39958;

DT 15-MAY-2002 (first entry)

XX Synthetic Gag polynucleotide sequence SEQ ID NO:21.

XX Human immunodeficiency virus type C; antigenic HIV type C protein;  
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
KW immunostimulant; gene therapy; gene; ds.

XX Human immunodeficiency virus; type C.  
OS Synthetic.

XX WO200204493-A2.

XX 17-JAN-2002.

XX 05-JUL-2001; 2001WO-US021241.

XX 05-JUL-2000; 2000US-00610313.

XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2002-154920/20.

XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunization or generation of packaging cell  
PT lines, particularly in gene therapy.

XX Example 1; Fig 6; 233pp; English.

XX The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (I). (I) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC Type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention

XX Sequence 1509 BP; 321 A; 560 C; 470 G; 158 T; 0 U; 0 Other;

Query Match 98.8%; Score 1491.4; DB 6; Length 1509;  
Best Local Similarity 99.3%; Pred. No. 7.7e-175;  
Matches 1498; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATGGGCGCCCGCGCCAGCATCTGCGCGCGAGAAAGTGGCAAGATCCGC 60  
Db 1 ATGGGCGCCCGCGCCAGCATCTGCGCGCGAGAAAGTGGCAAGATCCGC 60  
QY 61 CTGCGCCCGCGCGCAAGAACCTACATGCTGAAGCACTGTGTGGCGAGCCGCGAG 120  
Db 61 CTGCGCCCGCGCGCAAGAACCTACATGCTGAAGCACTGTGTGGCGAGCCGCGAG 120  
QY 121 CTGAGGGCTTCCGCTGAAACCCCGGCTGTGAGACCGCGAGGGCTGCAAGCAGATC 180  
Db 121 CTGAGGGCTTCCGCTGAAACCCCGGCTGTGAGACCGCGAGGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGACAGCCCGCTGACAGCCGAGCAACGAGAGCTGCGACCTGTACAAC 240  
Db 181 ATGAAGCAGCTGACAGCCCGCTGACAGCCGAGCAACGAGAGCTGCGACCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGACAGCGCGCATCGAGTCCGCAACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTACTGCGTGACAGCGCGCATCGAGTCCGCAACCAAGAGGCC 300  
QY 301 CTGACCAAGATCGAGAGAGCAGAACCAAGTCCAGCAGAAACCCAGCGCCCAAGAG 360  
Db 301 CTGACCAAGATCGAGAGAGCAGAACCAAGTCCAGCAGAAACCCAGCGCCCAAGAG 360  
QY 361 GCCGACGCGAAGGTGAGCGCAACTACCCCATCGTGCAAGACTTGACGGCCAGATGTTG 420  
Db 361 GCCGACGCGAAGGTGAGCGCAACTACCCCATCGTGCAAGACTTGACGGCCAGATGTTG 420  
QY 421 CACCAGGCCATCAGCCCCCGCACTGAAAGCTGTGGTGAAGTATCGAAGAGGCC 480  
Db 421 CACCAGGCCATCAGCCCCCGCACTGAAAGCTGTGGTGAAGTATCGAAGAGGCC 480  
QY 481 TTCAGCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGGAGGGCGCCACCCCGAGAG 540  
Db 481 TTCAGCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGGAGGGCGCCACCCCGAGAG 540  
QY 541 CTGAACACGATGTTGAACACCGTGGCGGCCACAGCGCCATGACATGCTGAAGGAC 600  
Db 541 CTGAACACGATGTTGAACACCGTGGCGGCCACAGCGCCATGACATGCTGAAGGAC 600  
QY 601 ACCATCAACGAGAGGCGCGGAGTGGGACCGCTGACCCCGTGACGGCGGCCCGCTG 660  
Db 601 ACCATCAACGAGAGGCGCGGAGTGGGACCGCTGACCCCGTGACGGCGGCCCGCTG 660  
QY 661 GCCCGCGCGAGATGCGCGACCCCGCGGAGCGACATGCGCGGCCACAGCACCTTG 720  
Db 661 GCCCGCGCGAGATGCGCGACCCCGCGGAGCGACATGCGCGGCCACAGCACCTTG 720  
QY 721 CAGGACGATCGCTGATGACCAACCCCGCTGCGCGCTGCGGCGACATCTACAAG 780  
Db 721 CAGGACGATCGCTGATGACCAACCCCGCTGCGCGCTGCGGCGACATCTACAAG 780  
QY 781 CGGTGATCATCTGAGGCTGAAACAGATCGTGGATGTACAGCCCGTGAGCATCTG 840  
Db 781 CGGTGATCATCTGAGGCTGAAACAGATCGTGGATGTACAGCCCGTGAGCATCTG 840  
QY 841 GACATCCGCGAGGCGCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAGACC 900  
Db 841 GACATCCGCGAGGCGCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAGACC 900  
QY 901 CTGCGCGCGAGCAGGCGCAACGAGAGCTGAAAGAACTGATGACCGAGACCTGTGTTG 960  
Db 901 CTGCGCGCGAGCAGGCGCAACGAGAGCTGAAAGAACTGATGACCGAGACCTGTGTTG 960  
QY 961 CAGAACGCGCAACCCGACATGCAAGACCAATCTGCGCGCTCTGCGCCCGCGCCACCTG 1020  
Db 961 CAGAACGCGCAACCCGACATGCAAGACCAATCTGCGCGCTCTGCGCCCGCGCCACCTG 1020  
QY 1021 GAGGAGATGATGACCGCTTGCCAGGGGCTGGCGCGCCCGGCCACCAAGGCCCGGTGCTG 1080

Db 1021 GAGAGATGATGACCGCTGCGCAGGGCGTGGGGCGCCCGGCCACAAAGGCCCGGTGCTG 1080  
QY 1081 GCCGAGCGATGAGCCAGGCCCAACAGCGTGAACATCATGTGACGAAGACAATTCAAG 1140  
Db 1081 GCCGAGCGCATGAGCCAGGCCCAACAGCGTGAACATCATGTGACGAAGACAATTCAAG 1140  
QY 1141 GGCCCCCGCGCAACGTGAAGTGTCTCACTGCGCAAGAGAGGCCACATCGCCAAGAAC 1200  
Db 1141 GGCCCCCGCGCAACGTGAAGTGTCTCACTGCGCAAGAGAGGCCACATCGCCAAGAAC 1200  
QY 1201 TGCCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGGCAAGAGGCCACCATGAAG 1260  
Db 1201 TGCCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGGCAAGAGGCCACCATGAAG 1260  
QY 1261 GACTGCAACCGAGCGCGCACTTCTGTGGGCAAGATCTGGCCCAAGCCACAAGGGCGCC 1320  
Db 1261 GACTGCAACCGAGCGCGCACTTCTGTGGGCAAGATCTGGCCCAAGCCACAAGGGCGCC 1320  
QY 1321 CCGGGAATTCTCTGCAAGACCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
Db 1321 CCGGGAATTCTCTGCAAGACCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
QY 1381 CCGGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCG 1440  
Db 1381 CCGGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCG 1440  
QY 1441 CGCGAGCCCTACCGCGAGCCCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
Db 1441 CGCGAGCCCTACCGCGAGCCCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
QY 1501 AGCCAGTAA 1509  
Db 1501 AGCCAGTAA 1509

RESULT 6  
ADM73763  
ID ADM73763 standard; DNA; 1509 BP.

XX AC ADM73763;  
XX DT 03-JUN-2004 (first entry)  
XX DE HIV-1 polynucleotide #6.  
XX KW HIV-1; gene; da; HIV pol; immune response; DNA immunisation;  
XX KW HIV type C protein; immunostimulant.  
XX OS Human immunodeficiency virus 1.  
XX PN US2003223961-A1.  
XX PD 04-DEC-2003.  
XX PF 05-JUL-2001; 2001US-00899575.  
XX PR 05-JUL-2000; 2000US-00610313.  
XX PA (MEGE/) MEGEDE J Z.  
XX PA (BARN/) BARNETT S W.  
XX PA (ENGE/) ENGELBRECHT S.  
XX PA (RENS/) RENSBURG B J V.  
XX PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg BJV;  
XX DR WPI; 2004-060515/06.  
XX PT New expression cassette comprising a polynucleotide sequence encoding an  
PT HIV pol polypeptide, useful in eliciting an immune response, in DNA  
PT immunization, generating of packaging cell lines or in producing HIV Type  
XX C proteins.

PS Example 1; SEQ ID NO 21; 160pp; English.  
XX The invention relates to an expression cassette comprising a  
CC polynucleotide sequence encoding an HIV pol polypeptide. The invention  
CC also relates to a recombinant expression system for use in a host cell  
CC comprising an expression cassette, where the polynucleotide sequence  
CC further comprises control elements capable of driving expression in the  
CC selected host cell, a cell comprising an expression cassette where the  
CC polynucleotide sequence further comprises control elements compatible  
CC with the expression in the cell and a composition for generating an  
CC immunological response, comprising an expression cassette. The expression  
CC cassette and the methods of the invention are useful in eliciting an  
CC immune response, in DNA immunisation, in generation of packaging cell  
CC lines and in producing HIV Type C proteins. This sequence represents an  
XX HIV-1 polynucleotide of the invention.  
SQ Sequence 1509 BP; 321 A; 560 C; 470 G; 158 T; 0 U; 0 Other;

Query Match 98.8%; Score 1491.4; DB 12; Length 1509;

Best Local Similarity 99.3%; Pred. No. 7.7e-175;

Matches 1498; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATGGCGCGCGCGCGCGCATCTCGCGCGCGCGAGAGAGCTGAGCAAGTGGAGAGATCCGC 60  
Db 1 ATGGCGCGCGCGCGCGCATCTCGCGCGCGCGAGAGAGCTGAGCAAGTGGAGAGATCCGC 60  
QY 61 CTGCGCGCGCGCGCGCAAGACATCATGTGAGACACCTGTGTGGCGCAGCCCGGAG 120  
Db 61 CTGCGCGCGCGCGCGCAAGACATCATGTGAGACACCTGTGTGGCGCAGCCCGGAG 120  
QY 121 CTGAGGGGCTTGGCCTTGAACCCCGGCTGTGAGAGACCGCGAGGGCTGCAAGCAGATC 180  
Db 121 CTGAGGGGCTTGGCCTTGAACCCCGGCTGTGAGAGACCGCGAGGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 181 ATGAAGCAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGCGCGCGCATTCGAGGTCCGCGACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTACTGCGTGCACGCGCGCGCATTCGAGGTCCGCGACCAAGAGGCC 300  
QY 301 CTGGAACAAGATCGAGAGAGAGAGACAAGTCCCGCAGAGAAGACCAGAGGCCAAGAG 360  
Db 301 CTGGAACAAGATCGAGAGAGAGAGACAAGTCCCGCAGAGAAGACCAGAGGCCAAGAG 360  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTAACCCATGTGAGAACTGCAAGGGCCAGATGTTG 420  
Db 361 GCCGACGGCAAGGTGAGCCAGAACTAACCCATGTGAGAACTGCAAGGGCCAGATGTTG 420  
QY 421 CACGAGGCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 421 CACGAGGCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
QY 481 TTCAAGCGCGGAGTATCCCATGTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Db 481 TTCAAGCGCGGAGTATCCCATGTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
QY 541 CTGAACACGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
Db 541 CTGAACACGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
QY 601 ACCATCAACGAGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
Db 601 ACCATCAACGAGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
QY 661 GCGCGCGCGCGCGAGTGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
Db 661 GCGCGCGCGCGCGAGTGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
QY 721 CAGGACGAGATCGCTGTGATGACCAAGACCCCGCGCGCGCGCGCGCGCGCGCGCG 780  
Db 721 CAGGACGAGATCGCTGTGATGACCAAGACCCCGCGCGCGCGCGCGCGCGCGCGCG 780

QY 781 CGGTGATCATCTGGGCTGAACAAGATGTCGGATGTAACAGCCCCGTGAGCATCTCTG 840  
DB 781 CGCTGATCATCTGGGCTGAACAAGATGTCGGATGTAACAGCCCCGTGAGCATCTCTG 840  
QY 841 GACATCCGCGCAGGCCCCCAAGAGAGCCCTCCGCGACTAGTGAGACCGCTTCTTCAAGACC 900  
DB 841 GACATCCGCGCAGGCCCCCAAGAGAGCCCTCCGCGACTAGTGAGACCGCTTCTTCAAGACC 900  
QY 901 CTGCGCGCCGACGAGCCCAACAGACGTGAAGAACTGGATGACCGAGACCCCTGCTGTG 960  
DB 901 CTGCGCGCCGACGAGCCCAACAGACGTGAAGAACTGGATGACCGAGACCCCTGCTGTG 960  
QY 961 CAGAACGCCAACCCCGACTGCAAGACCACTCTGCGGCTCTCGGCCCGCGCCACCTCTG 1020  
DB 961 CAGAACGCCAACCCCGACTGCAAGACCACTCTGCGGCTCTCGGCCCGCGCCACCTCTG 1020  
QY 1021 GAGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCCCGGCCCAAGAGCCCGCTGCTG 1080  
DB 1021 GAGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCCCGGCCCAAGAGCCCGCTGCTG 1080  
QY 1081 GCCGAGCGGATGAGCCAGGCCCAACAGCGTGAACATCATGATGACAGAGAGCAACTTCAAG 1140  
DB 1081 GCCGAGCGGATGAGCCAGGCCCAACAGCGTGAACATCATGATGACAGAGAGCAACTTCAAG 1140  
QY 1141 GGGCCCCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGCGCCACATCGCCAGAAC 1200  
DB 1141 GGGCCCCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGCGCCACATCGCCAGAAC 1200  
QY 1201 TGCCGCGCCCCCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGCCACCATGAAG 1260  
DB 1201 TGCCGCGCCCCCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGCCACCATGAAG 1260  
QY 1261 GACTGCAACGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGCCAGCCCAAGAGGCGCG 1320  
DB 1261 GACTGCAACGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGCCAGCCCAAGAGGCGCG 1320  
QY 1321 CCGGCAACTTCTGCGAGAACCGCAGCGAGCGCGCGCGCCCGCCACCGTGCACCGCCCC 1380  
DB 1321 CCGGCAACTTCTGCGAGAACCGCAGCGAGCGCGCGCGCGCCCGCCACCGTGCACCGCCCC 1380  
QY 1381 CCGGCGAGAGCTTCCGCTTGAAGAGAACACCGCGCGCGCGCCCGCCAGAGAGCGCAAGAG 1440  
DB 1381 CCGGCGAGAGCTTCCGCTTGAAGAGAACACCGCGCGCGCGCGCCCGCCAGAGAGCGCAAGAG 1440  
QY 1441 CGGAGCCCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
DB 1441 CGGAGCCCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
QY 1501 AGCCAGTAA 1509  
DB 1501 AGCCAGTAA 1509

RESULT 7  
ID AA51626  
AA51626 standard; DNA; 1509 BP.

AC AA51626;  
DT 31-OCT-2000 (first entry)  
DE HIV codon-optimized synthetic Gag polynucleotide.  
KW Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;  
KW DNA immunization; packaging cell line; antigen presentation; ss.  
OS Human immunodeficiency virus; type C strain AF110967.  
OS Synthetic.  
XX WO200039304-A2.  
XX 06-JUL-2000.

XX 30-DEC-1999; 99WO-US031273.  
PR 31-DEC-1998; 98US-011495P.  
PR 01-SEP-1999; 99US-0152195P.  
XX (CHIR ) CHIRON CORP.  
PI Barnett S, Zur Megede J;  
XX WPI; 2000-452401/39.  
PT Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env  
PT polypeptide and the polypeptide useful for immunizing a mammal especially  
PT human against HIV.  
XX  
PS Disclosure; Page 104; 113p; English.  
CC Expression cassettes comprising a polynucleotide encoding antigenic type  
CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
CC in DNA immunization, generation of packaging cell lines and production of  
CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
CC cassettes exhibit increased potency for induction of cytotoxic T-  
CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-  
CC assemble into non-infectious virus-like particles which are used as a  
CC matrix for the proper presentation of an antigen entrapped or associated  
CC to the immune system of the host  
XX  
SQ Sequence 1509 BP; 321 A; 559 C; 471 G; 158 T; 0 U; 0 Other;

Query Match 98.7%; Score 1489.8; DB 3; Length 1509;  
Best Local Similarity 99.2%; Pred. No. 1.2e-174;  
Matches 1497; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGGGCGCCGCGCGCAGCATCTGCGCGCGAGAACTGCAAGTGGGAGAAAGATCCGC 60  
DB 1 ATGGGCGCCGCGCGCAGCATCTGCGCGCGAGAACTGCAAGTGGGAGAAAGATCCGC 60  
QY 61 CTGCGCCCGCGCGCAAGAGCACTACTAGTCTGAAGCACTGTGTGGGCCAGCGCGAG 120  
DB 61 CTGCGCCCGCGCGCAAGAGCACTACTAGTCTGAAGCACTGTGTGGGCCAGCGCGAG 120  
QY 121 CTGAGGGCTTGCCTTGAACCCCGGCTGTGAGACCGCGAGGCTGCAAGCAGATC 180  
DB 121 CTGAGGGCTTGCCTTGAACCCCGGCTGTGAGACCGCGAGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGACGCGCGCTGTGACAGCCGACCGAGAGCTGCGCAGCTGTACAAC 240  
DB 181 ATGAAGCAGCTGACGCGCGCTGTGACAGCCGACCGAGAGCTGCGCAGCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGTCGTGACGCGCGCATGAGTCCGCGACCAAGAGAGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGTCGTGACGCGCGCATGAGTCCGCGACCAAGAGAGGCC 300  
QY 301 CTGACACAGATCGAGGAGAGAGCAAGTCCACAGAGAAAGACCAAGAGAGAGAG 360  
DB 301 CTGACACAGATCGAGGAGAGAGCAAGTCCACAGAGAAAGACCAAGAGAGAGAG 360  
QY 361 GCCGACGCAAGGTGAGCCAGAACTAACCCCATCTGTGCAAGAACTGCAAGGCGCAATGGTG 420  
DB 361 GCCGACGCAAGGTGAGCCAGAACTAACCCCATCTGTGCAAGAACTGCAAGGCGCAATGGTG 420  
QY 421 CACGAGCCATCAGCCCCCGCACCTGAAAGCTGTGTAAGTGTGAGAGAGAGGCC 480  
DB 421 CACGAGCCATCAGCCCCCGCACCTGAAAGCTGTGTAAGTGTGAGAGAGAGGCC 480  
QY 481 TTCAAGCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCGAGAG 540  
DB 481 TTCAAGCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCGAGAG 540  
QY 541 CTGAACAGATGTTGAACACCGTGGCGCGCCACAGCGCGCCATGATGCTGAAGAG 600  
DB 541 CTGAACAGATGTTGAACACCGTGGCGCGCCACAGCGCGCCATGATGCTGAAGAG 600

```

QY 601 ACCATCAAGAGAGGCGCCGAGTGGACCGCTGCACCCCGTGAGAGCGCCCGTG 660
   |||||||
Db 601 ACCATCAAGAGAGGCGCCGAGTGGACCGCTGCACCCCGTGAGAGCGCCCGTG 660
QY 661 GCGCGCGCGAGATGCGCGGACCCCGCGGAGGAGCATCGCGCGCGGAGCAGACCTG 720
   |||||||
Db 661 GCGCGCGCGAGATGCGCGGACCCCGCGGAGGAGCATCGCGCGCGGAGCAGACCTG 720
QY 721 CAGAGCAGATCGCTGGATGACAGCAACCCCGCTGCGCGGAGCATCTACAG 780
   |||||||
Db 721 CAGAGCAGATCGCTGGATGACAGCAACCCCGCTGCGCGGAGCATCTACAG 780
QY 781 CGGTGATCATCTGGGCTGACACAAGATCGTGCGATGTACAGCCCGTGAGCATCTG 840
   |||||||
Db 781 CGGTGATCATCTGGGCTGACACAAGATCGTGCGATGTACAGCCCGTGAGCATCTG 840
QY 841 GACATCCGCGCAGGCGCCCAAGAGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGAC 900
   |||||||
Db 841 GACATCCGCGCAGGCGCCCAAGAGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGAC 900
QY 901 CTGCGCGCGGAGCAGGCGCACCCAGAGCGTGAAGACTGTGATGACCGAGACCTGCTG 960
   |||||||
Db 901 CTGCGCGCGGAGCAGGCGCACCCAGAGCGTGAAGACTGTGATGACCGAGACCTGCTG 960
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCGCGCGGAGCAGCCTG 1020
   |||||||
Db 961 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCGCGCGGAGCAGCCTG 1020
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGTGGGCGCGCGCGGAGCAGGCGCGCTG 1080
   |||||||
Db 1021 GAGGAGATGATGACCGCTGCGCAGGCGTGGGCGCGCGCGGAGCAGGCGCGCTG 1080
QY 1081 GCGGAGCGGATGAGCGGCGGAGCAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140
   |||||||
Db 1081 GCGGAGCGGATGAGCGGCGGAGCAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140
QY 1141 GCGCGCGCGGAGCAGCGTGAAGTGTCTCAACTGCGGAGAGGCGGAGCAGTGGCAAGAAC 1200
   |||||||
Db 1141 GCGCGCGCGGAGCAGCGTGAAGTGTCTCAACTGCGGAGAGGCGGAGCAGTGGCAAGAAC 1200
QY 1201 TGCGCGCGCGCGGAGAGAGGCTGTCTGGAAGTGGGAGAGGCGGAGCAGATGAAG 1260
   |||||||
Db 1201 TGCGCGCGCGCGGAGAGAGGCTGTCTGGAAGTGGGAGAGGCGGAGCAGATGAAG 1260
QY 1261 GACTGCAACGAGCGCGGAGCGCAACTTCTGGGCAAGATCTGCGGAGCAGCAGGCGCGC 1320
   |||||||
Db 1261 GACTGCAACGAGCGCGGAGCGCAACTTCTGGGCAAGATCTGCGGAGCAGCAGGCGCGC 1320
QY 1321 CCGGCAACTTCTGAGAACCGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
   |||||||
Db 1321 CCGGCAACTTCTGAGAACCGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
QY 1381 CCGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGGAGAGGCGGAGGCGGAGGAG 1440
   |||||||
Db 1381 CCGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGGAGAGGCGGAGGCGGAGGAG 1440
QY 1441 CGGAGCGCTACCGCGAGCGCGCTGACCGCGCTGCGGAGCGCTTGGAGCGGCGCGCTG 1500
   |||||||
Db 1441 CGGAGCGCTACCGCGAGCGCGCTGACCGCGCTGCGGAGCGCTTGGAGCGGCGCGCTG 1500
QY 1501 AGCCAGTAA 1509
   |||||||
Db 1501 AGCCAGTAA 1509

```

```

DT 08-NOV-2002 (first entry)
XX
DE HIV-1 p55gag polypeptide coding sequence 4.
XX
KW HIV; ds; vaccine; gene; immune response; microparticle;
KW adsorbent surface; poly(alpha-hydroxy acid); polyhydroxy butyric acid;
KW polycaprolactone; polyorthoester; polycyanoacrylate; detergent;
KW submicron emulsion; viral infection; bacterial infection;
KW parasitic infection; HIV-1 p55gag polypeptide.
XX
OS Human immunodeficiency virus 1.
XX
FH Key Location/Qualifiers
FT variation replace(282, C)
FT /*tag= a
FT variation replace(331, T)
FT /*tag= b
FT variation replace(332, C)
FT /*tag= c
FT variation replace(549, G)
FT /*tag= d
FT variation replace(553, T)
FT /*tag= e
FT variation replace(783, G)
FT /*tag= f
FT variation replace(816, G)
FT /*tag= g
FT variation replace(999, T)
FT /*tag= h
FT variation replace(1002, C)
FT /*tag= i
FT variation replace(1089, G)
FT /*tag= j
FT variation replace(1149, G)
FT /*tag= k
FT variation replace(1158, C)
FT /*tag= l
XX
PN WO200226209-A2.
XX
PD 04-APR-2002.
XX
PF 28-SEP-2001; 2001WO-US030540.
XX
PR 28-SEP-2000; 2000US-0236105P.
PR 30-AUG-2001; 2001US-0315905P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI O'hagan D, Otten G, Donnelly JJ, Polo JM, Barnett S, Singh M,
PI Ulmer J, Dubensky TW;
XX
DR WPI; 2002-519084/55.
XX
PT A microparticle to which a biologically active macromolecule is adsorbed,
PT for use as a vaccine composition to treat viral, bacterial or parasitic
PT infections, comprises a polymer microparticle, a detergent and a
PT submicron emulsion.
XX
PS Claim 72; Fig 6; 100pp; English.
XX
CC The invention relates to a method of raising an immune response in a host
CC animal. The method of the invention comprises administering a
CC microparticle that has an adsorbent surface to which a first biologically
CC active macromolecule (e.g. a nucleic acid) has been adsorbed. The
CC microparticle comprises a polymer microparticle of poly(alpha-hydroxy
CC acid), a polyhydroxy butyric acid, a polycaprolactone, a polyorthoester,
CC a polycyanoacrylate, a detergent, and submicron emulsion. The method/
CC microparticle of the invention is useful for immunising a host animal
CC against viral, bacterial or parasitic infections. The present DNA
CC sequence encodes a HIV-1 p55gag polypeptide. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX

```

RESULT 8  
 AAL44553  
 ID AAL44553 standard; DNA; 1509 BP.  
 XX  
 AC AAL44553;  
 XX  
 DT 29-AUG-2003 (revised)

Sequence 1509 BP; 321 A; 559 C; 471 G; 158 T; 0 U; 0 Other;

Query Match	98.7%;	Score 1489.8;	DB 6;	Length 1509;
Best Local Similarity	99.2%;	Pred. No. 1.2e-174;		
Matches 1497;	Conservative	0;	Mismatches 12;	Indels 0;
			Gaps	0;

QY	1	ATGGCGCCCGCCCAAGCATCTCTGCGCGCGAGAAAGCTGGACAAGTGGAGAAATCCGC	60
Db	1	ATGGCGCGCCCGCCCAAGCATCTCTGCGCGCGAGAAAGCTGGACAAGTGGAGAAATCCGC	60
QY	61	CTGCGCCCCGGCGGCAAGAAAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG	120
Db	61	CTGCGCCCCGGCGGCAAGAAAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG	120
QY	121	CTGGAGGGCTTGCCCTTGAAACCCCGGCTGTGGAGACCGCCGAGGGCTGCAAGCATC	180
Db	121	CTGGAGGGCTTGCCCTTGAAACCCCGGCTGTGGAGACCGCCGAGGGCTGCAAGCATC	180
QY	181	ATGAAGCAGCTGCAGACCCGCGCTGCAGACCGGCAACGAGAGCTGCAGACCTGTACAAC	240
Db	181	ATGAAGCAGCTGCAGACCCGCGCTGCAGACCGGCAACGAGAGCTGCAGACCTGTACAAC	240
QY	241	ACCGTGGCCACCTGTACTGCGTGCACGCGCGCATCGAGTCCGCGACACCAAGAGGCC	300
Db	241	ACCGTGGCCACCTGTACTGCGTGCACGCGCGCATCGAGTCCGCGACACCAAGAGGCC	300
QY	301	CTGGACAAGATCGAGGAGGAGCAAAAGTCCAGCAGAAAGACCAGAGGCCAAGAG	360
Db	301	CTGGACAAGATCGAGGAGGAGCAAAAGTCCAGCAGAAAGACCAGAGGCCAAGAG	360
QY	361	GCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACTTGACAGGCGCAGATG	420
Db	361	GCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACTTGACAGGCGCAGATG	420
QY	421	CACCAGGCCATCAGCCCCCGCAACCTGAAGCGCTGGGTGAAGGTGATCGAGGAAAGGCC	480
Db	421	CACCAGGCCATCAGCCCCCGCAACCTGAAGCGCTGGGTGAAGGTGATCGAGGAAAGGCC	480
QY	481	TTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTTGAGCGAGGGCGCAACCCCGAGAC	540
Db	481	TTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTTGAGCGAGGGCGCAACCCCGAGAC	540
QY	541	CTGAACACGATGTTGAACACCGTGGCGGCGCAACAGCGCGCATGCAGATGCTGAAGAC	600
Db	541	CTGAACACGATGTTGAACACCGTGGCGGCGCAACAGCGCGCATGCAGATGCTGAAGAC	600
QY	601	ACCATCAACGAGGAGCGCGCGAGTGGACCGCCTTGACCCCGTGCAAGCGCGCCCGTG	660
Db	601	ACCATCAACGAGGAGCGCGCGAGTGGACCGCCTTGACCCCGTGCAAGCGCGCCCGTG	660
QY	661	GCCCCCGGCAGATGCGCGAACCCCGCGCGAGCAGATCGCGCGGCCCAACGACCCCTG	720
Db	661	GCCCCCGGCAGATGCGCGAACCCCGCGCGAGCAGATCGCGCGGCCCAACGACCCCTG	720
QY	721	CAGAGCAGATGCGCTGATGACCAAGACCCCCCGTGCCTGGCGGCAATCTACAAG	780
Db	721	CAGAGCAGATGCGCTGATGACCAAGACCCCCCGTGCCTGGCGGCAATCTACAAG	780
QY	781	CGGTGATCATCTGCGCTGAACAAGATCGTGCAGATGTAACGCCCCGTGAGCATCCTG	840
Db	781	CGGTGATCATCTGCGCTGAACAAGATCGTGCAGATGTAACGCCCCGTGAGCATCCTG	840
QY	841	GACATCCGCGCAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGACC	900
Db	841	GACATCCGCGCAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGACC	900
QY	901	CTGCGCGCGGAGCAGGCGCAACCAAGACGTGAAGAACTGATGACCGAGACCTTGCTG	960
Db	901	CTGCGCGCGGAGCAGGCGCAACCAAGACGTGAAGAACTGATGACCGAGACCTTGCTG	960
QY	961	CAGAACGCCAACCCCGACTGCAAGAACCATCTGCGCGCTTCCGGCCCCCGGCGCCACCTG	1020
Db	961	CAGAACGCCAACCCCGACTGCAAGAACCATCTGCGCGCTTCCGGCCCCCGGCGCCACCTG	1020

QY	1021	GAGGAGATGATGA	CCGCTCTGCCAGGGCGCTGGGCGGCCCCGGGCCACAAAGCCCGCTGCTG	1080
Db	1021	GAGGAGATGATGA	CCGCTCTGCCAGGGCGCTGGGCGGCCCCGGGCCACAAAGCCCGCTGCTG	1080
QY	1081	GCCGAGGCGGATGAG	CCAGGCCCAACAGCGTGAACATCATGTATGCAGAAAGCAACTTCAAG	1140
Db	1081	GCCGAGGCGGATGAG	CCAGGCCCAACAGCGTGAACATCATGTATGCAGAAAGCAACTTCAAG	1140
QY	1141	GGCCCCCGGCGCAACGTCA	AGTGTCTTCACTGCGGCAAGAGAGGGCCACATCGCCCAAGAAC	1200
Db	1141	GGCCCCCGGCGCAACGTCA	AGTGTCTTCACTGCGGCAAGAGAGGGCCACATCGCCCAAGAAC	1200
QY	1201	TGCGCGCGCCCCCGCAAGAA	GGGCTGTGGAAGTGCGGCAAGAGAGGGCCACAGATGAAG	1260
Db	1201	TGCGCGCGCCCCCGCAAGAA	GGGCTGTGGAAGTGCGGCAAGAGAGGGCCACAGATGAAG	1260
QY	1261	GACTGCACCGGAGCGCCAG	GGCCAACTTCTGGGGCAAGATGTGGCCACAGCCACAAGGGCGCG	1320
Db	1261	GACTGCACCGGAGCGCCAG	GGCCAACTTCTGGGGCAAGATGTGGCCACAGCCACAAGGGCGCG	1320
QY	1321	CCCGGCAACTTCTTGCA	GAAACCGCAGCGAGCCGCGCGGCCCCCAACCGTGCCACCGGCCCC	1380
Db	1321	CCCGGCAACTTCTTGCA	GAAACCGCAGCGAGCCGCGCGGCCCCCAACCGTGCCACCGGCCCC	1380
QY	1381	CCCGCGGAGAGCTTCCG	CTTGCAGAGAGACCAACCCCGCGCCCAAGCAGAGAGCCCAAGGAC	1440
Db	1381	CCCGCGGAGAGCTTCCG	CTTGCAGAGAGACCAACCCCGCGCCCAAGCAGAGAGCCCAAGGAC	1440
QY	1441	CGCGAGCCCTTACCGCG	AGCCCTGACCGCCCTGCGCAGCCTGTTCCGCAAGCGGCCCCCTTG	1500
Db	1441	CGCGAGCCCTTACCGCG	AGCCCTGACCGCCCTGCGCAGCCTGTTCCGCAAGCGGCCCCCTTG	1500
QY	1501	AGCCAGTTAA	1509	
Db	1501	AGCCAGTTAA	1509	

```

RESULT 9
ABL39972
ID ABL39972 standard; DNA; 1494 BP.
XX
XX ABL39972;
XX
XX 15-MAY-2002 (first entry)
XX
XX Synthetic Gag polynucleotide sequence SEQ ID NO:51.
DE
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
XX Human immunodeficiency virus; type C.
OS Synthetic.
XX
XX WO200204493-A2.
PN
XX
XX 17-JAN-2002.
PD
XX
XX 05-JUL-2001; 2001WO-US021241.
PF
XX
XX 05-JUL-2000; 2000US-00610313.
PR
XX
XX (CHIR ) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
XX WPI; 2002-154920/20.
XX
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
PT

```

XX Claim 7; Fig 22; 233bp; English.  
PS  
XX The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (1). (I) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 1494 BP; 325 A; 557 C; 455 G; 157 T; 0 U; 0 Other;

Query Match 86.0%; Score 1297.6; DB 6; Length 1494;  
Best Local Similarity 92.7%; Pred. No. 4.7e-151;  
Matches 1402; Conservative 0; Mismatches 89; Indels 21; Gaps 3;

QY 1 ATGGGCGCCCGCCGACATCTCGCGCGGAGAAAGCTGGACAAGTGGAGAAAGATCCGC 60  
Db 1 ATGGGCGCCCGCCGACATCTGAGCGCGGAGAAAGCTGGACAAGTGGAGAGCGCATCCGC 60  
QY 61 CTGCGCGCCCGCGCGGCAAGAACACTACATGCTGAAGCACTGTGTGGGCCAGCCGCGAG 120  
Db 61 CTGCGCGCCCGCGCGGCAAGAACACTACATGCTGAAGCACTGTGTGGGCCAGCCGCGAG 120  
QY 121 CTGAGAGGCTTCGCGCTGAAACCCCGGCTGTGAGACCGCGCGAGGCTGCAAGCAGATC 180  
Db 121 CTGAGAGGCTTCGCGCTGAAACCCCGGCTGTGAGACCGAGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGCAGCGCGCGCTGCAAGACCGGACCGAGAGAGCTGCGCAGCTGTACAAC 240  
Db 181 ATGAAGCAGCTGCAGCGCGCGCTGCAAGACCGGACCGAGAGAGCTGCGCAGCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCTGTCAGCGCGCGCATGAGGTCCGCGACACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTACTGCTGTCAGCGCGCATGAGGTCCGCGACACCAAGAGGCC 300  
QY 301 CTGGAACAAGATCGAGAGAGAGCAGAACAGTCCAGCAGAAAGACCCAGGCGCCAAAGGCC 360  
Db 301 CTGGAACAAGATCGAGAGAGAGCAGAACAGTCCAGCAGAAAGAGCCCAAGGCGCCAAAGGCC 360  
QY 361 GCCACGCGCAAGGTGAGCCAGAACCTACCTGTCAGAACTGTGAGAGGCGCAGATGGTG 420  
Db 361 GCCACGCGCAAGGTGAGCCAGAACCTACCTGTCAGAACTGTGAGAGGCGCAGATGGTG 420  
QY 421 CACCAAGCCATCAGCCCCCGCACCTGAAAGCGCTGATCAAGGTGATCGAGAGAGGCC 480  
Db 421 CACCAAGCCATCAGCCCCCGCACCTGAAAGCGCTGATCAAGGTGATCGAGAGAGGCC 480  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGGCCACCCCAAGAC 540  
Db 481 TTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGGCCACCCCAAGAC 540  
QY 541 CTGAACACGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGAGATGCTGAAGGAC 600  
Db 541 CTGAACACGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGAGATGCTGAAGGAC 600  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGGAGACCGCTGCAACCCGCTGAGCGCGCGCTG 660  
Db 601 ACCATCAACGAGAGGCGCGCGAGTGGGAGACCGCACCAACCCGCTGAGCGCGCGCTG 660  
QY 661 GCGCGCGCGCAGATGCGCGAGACCGCGCGCGAGCAGCATGCGCGGCAACCAAGACCTG 720  
Db 661 GCGCGCGCGCAGATGCGCGAGACCGCGCGCGAGCAGCATGCGCGGCAACCAAGACCTG 720  
QY 721 CAGGAGCAGATGCGCTGATGACCAAGCAACCCCGCTGCGGTGGCGACATCTACAAG 780  
Db 721 CAGGAGCAGATGCGCTGATGACCAAGCAACCCCGCTGCGGTGGAGACATCTACAAG 780

QY 781 CGGTGATCATCTGGGCGCTGAACAGATCGTGGGATGTACAGCCCCGTGAGCATCTG 840  
Db 781 CGGTGATCATCTGGGCGCTGAACAGATCGTGGGATGTACAGCCCCGTGAGCATCTG 840  
QY 841 GACATCCGCGAGGCGCCCAAGAGGCTTCGCGCACTACGTGAGCGCTTCTTAAGACC 900  
Db 841 GACATCAAGCAGGCGCCCAAGAGGCTTCGCGCACTACGTGAGCGCTTCTTAAGACC 900  
QY 901 CTGGCGCGCGAGCAGGCGCACCGAGAGCTGAAGACTGATGACCGAGACCTGTGCTG 960  
Db 901 CTGGCGCGCGAGCAGGCGCACCGAGAGCTGAAGACTGATGACCGAGACCTGTGCTG 960  
QY 961 CAGAACGCCAACCCGACTGCAAGACCATCTGCGGCTCTGCGCCCGCGCGCACCTG 1020  
Db 961 CAGAACGCCAACCCGACTGCAAGACCATCTGCGGCGCTGCGCCCGCGCGCACCTG 1020  
QY 1021 GAGGAGATGATGACCGCTGCGAGGCGTGGCGCGCCCGCGCAAGGCGCGCTGCTG 1080  
Db 1021 GAGGAGATGATGACCGCTGCGAGGCGTGGCGCGCCCGCGCAAGGCGCGCTGCTG 1080  
QY 1081 GCCGAGCGGATGAGCCAGGCGCAAGCGTGAACATCATGATGACGAAGCAACTTCAAG 1140  
Db 1081 GCCGAGCGGATGAGCCAGGCGCAAGCGTGAACATCATGATGACGAAGCAACTTCAAG 1137  
QY 1141 GCGCGCGCGCGCAACGTCAGTGTCTCAACTGCGCGCAAGAGGCGCATGCGCAAGAAC 1200  
Db 1138 GCGAGCAACCGCATCATCAAGTGTCTCAACTGCGCGCAAGAGTGGGCAATGCGCGCAAC 1197  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGGCAAGAGGCGCACCAAGTGAAG 1260  
Db 1198 TGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGGCAAGAGGCGCACCAAGTGAAG 1257  
QY 1261 GACTGCACCGAGGCGCAAGGCTCTGCGGCAAGATCTGGCGAGCCACCAAGGCGCGC 1320  
Db 1258 GACTGCACCGAGGCGCAAGGCTCTGCGGCAAGATCTGGCGAGCCACCAAGGCGCGC 1317  
QY 1321 CCGGCAACTTCTGCAAGAACCGAGCGAGC---CGCGCGCGCGCACCGTGCCACCGCC 1377  
Db 1318 CCGGCAACTTCTGCAAGAACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1377  
QY 1378 CCGCGCGCGGAGGCTTCGCTGAGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCG 1437  
Db 1378 CCGCGCGCGGAGGCTTCGCTGAGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCG 1437  
QY 1438 GACCGCGAGCCCTACCGCGAGCGCGCTGACCGCGCTGCGAGCTGTTCGCGAGCGCGCC 1497  
Db 1438 -----CGGAGCGCGCTGACCGCGCTGAGAGAGCGCTGTTCGCGAGCGAGCGCC 1482  
QY 1498 CTGAGCCAGTAA 1509  
Db 1483 CTGAGCCAGTAA 1494

RESULT 10  
ADM73785  
ID ADM73785 standard; DNA; 1494 BP.  
XX  
AC ADM73785;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE HIV-1 polynucleotide #20.  
XX  
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
KW HIV type C protein; immunostimulant.  
XX  
OS Human immunodeficiency virus 1.  
PN US2003223961-A1.  
XX  
PD 04-DEC-2003.  
XX  
PF 05-JUL-2001; 2001US-00899575.

XX 05-JUL-2000; 2000US-00610313.  
PR  
XX  
PA (MEGE/) MEGEDE J Z.  
PA (BARN/) BARNETT S W.  
PA (ENGE/) ENGELBRECHT S.  
PA (RENS/) RENSBURG E J V.  
XX  
PI Mege de JZ, Barnett SW, Engelbrecht S, Rensburg EJV;  
XX  
XX WPI; 2004-060515/06.  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding an  
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
PT immunization, generating of packaging cell lines or in producing HIV Type  
PT C proteins.  
XX  
XX Claim 7; SEQ ID NO 51; 160pp; English.  
XX  
CC The invention relates to an expression cassette comprising a  
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
CC also relates to a recombinant expression system for use in a host cell  
CC comprising an expression cassette, where the polynucleotide sequence  
CC further comprises control elements capable of driving expression in the  
CC selected host cell, a cell comprising an expression cassette where the  
CC polynucleotide sequence further comprises control elements compatible  
CC with the expression in the cell and a composition for generating an  
CC immunological response, comprising an expression cassette. The expression  
CC cassette and the methods of the invention are useful in eliciting an  
CC immune response, in DNA immunisation, in generation of packaging cell  
CC lines and in producing HIV Type C proteins. This sequence represents an  
CC HIV-1 polynucleotide of the invention.  
XX  
SQ Sequence 1494 BP; 325 A; 557 C; 455 G; 157 T; 0 U; 0 Other;  
  
Query Match 86.0%; Score 1297.6; DB 12; Length 1494;  
Best Local Similarity 92.7%; Pred. No. 4.7e-151;  
Matches 1402; Conservative 0; Mismatches 89; Indels 21; Gaps 3;  
  
QY 1 ATGGGCGCCCGCGCCAGCATCTCGCGCGGCGAGAGCTGGACAAGTGGAGAGAATCCGC 60  
DB 1 ATGGGCGCCCGCGCCAGCATCTCGAGCGGCGGAGCTGGACAAGTGGAGAGCGCATCCGC 60  
  
QY 61 CTGCGCCCCGCGCGCAAGAGCACTAATGCTGAAGCACTGTGTGGGCGGCGGAG 120  
DB 61 CTGCGCCCCGCGCGCAAGAGCACTAATGCTGAAGCACTGTGTGGGCGGCGGAG 120  
  
QY 121 CTGAGGCGCTTCCGCTGAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCATC 180  
DB 121 CTGAGGCGCTTCCGCTGAACCCCGGCTGTGAGACCGAGGAGGCTGCAAGCATC 180  
  
QY 181 ATGAAGCAGCTGACGCCCGCTGAGACCGGACCGAGAGCTGGCGAGCTGTACAAC 240  
DB 181 ATGAAGCAGCTGACGCCCGCTGAGACCGGACCGAGAGCTGGCGAGCTGTACAAC 240  
  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGCGGCAATCGAGTCCGCGACCAAGAGAGCC 300  
DB 241 ACCGTGGCCACCTGTACTGCGTGCACAGGGGCAATCGAGTGGCGGACCAAGAGAGCC 300  
  
QY 301 CTGGAACAAGATCAGAGAGAGAGCAACAAGTCCAGAGAGAGCCAGAGGCAAGAG 360  
DB 301 CTGGAACAAGATCAGAGAGAGAGCAACAAGTCCAGAGAGAGCCAGAGGCAAGAG 360  
  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATGTGCAAGACCTGCAAGGCGCATGTG 420  
DB 361 GCCGACGGCAAGGTGAGCCAGAACTACCCCATGTGCAAGACGCCCAAGGCGCATGTG 420  
  
QY 421 CACCAAGGCATCAGCCCCCGCACTTGAAGCGCTGGGTGAAGTGTATGAGAGAGGCC 480  
DB 421 CACCAAGGCATCAGCCCCCGCACTTGAAGCGCTGGATCAAGTGTATGAGAGAGGCC 480  
  
QY 481 TTCAGCCCCGAGGTGATCCCAATGTTCACCGGCTGAGCGAGGGCGCCACCCCGAGAC 540  
DB 481 TTCAGCCCCGAGGTGATCCCAATGTTCACCGGCTGAGCGAGGGCGCCACCCCGAGAC 540

DB 481 TTCACCCCGAGAGATCCCAATGTTCACCGGCTGAGCGAGGGCGCCACCCCGAGAG 540  
QY 541 CTGAACACGATGTTGAACACCGTGGGGCGGCCCAACAGGCGCCCATGCAATGCTGAAGAG 600  
DB 541 CTGAACACGATGCTGAACACCGTGGGGCGGCCCAACAGGCGCCCATGCAATGCTGAAGAG 600  
QY 601 ACCATCAACGAGAGGGCGCCCGAGTGGAGCCGCTGCACCCCGTGCAGGCGGCGCCG 660  
DB 601 ACCATCAACGAGAGGGCGCCCGAGTGGAGCCGCGACCCACCCCGTGCAGGCGGCGCCG 660  
QY 661 GCGCCGCGCAGATGCGCGGACCCCGCGGCGAGCGACATCGCGCGGCGCCACCAAGCC 720  
DB 661 GCGCCGCGCAGATGCGCGGAGCGGCGCGGCGGCGAGCATCGCGGCGCCACCAAGCC 720  
QY 721 CAGGACGAGATCGCTGATGACCAAGCAACCCCGGCGGCGGCGGCGGCGGCGGCGG 780  
DB 721 CAGGACGAGATCGCTGATGACCAAGCAACCCCGGCGGCGGCGGCGGCGGCGGCGG 780  
QY 781 CGGTGATCATCTGGGCGCTGAACAAGATGTGCGGATGTACAGCCCGGTGAGCATCTG 840  
DB 781 CGGTGATCATCTGGGCGCTGAACAAGATGTGCGGATGTACAGCCCGGTGAGCATCTG 840  
QY 841 GACATCCGCGAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGGAACCGCTTCTCAAG 900  
DB 841 GACATCAAGAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGGAACCGCTTCTCAAG 900  
QY 901 CTGCGCGCGGAGCAGGCGCAACCAAGACGTGAAGAACTGATGACCGAGACCCCTGTG 960  
DB 901 CTGCGCGCGGAGCAGGCGCAACCAAGACGTGAAGAACTGATGACCGAGACCCCTGTG 960  
QY 961 CAGAACGCAACCCCGACTGCAAGACCATCTGCGGCTCTGCGCCCGCGCGCACCTG 1020  
DB 961 CAGAACGCAACCCCGACTGCAAGACCATCTGCGGCTCTGCGCCCGCGCGCACCTG 1020  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1080  
DB 1021 GAGGAGATGATGACCGCTGCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1080  
QY 1081 GCGGAGCGATGAGCGGCGCAACAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
DB 1081 GCGGAGCGATGAGCGGCGCAACAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
QY 1141 GCGCCCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGCGGCGGCGGCGGCG 1200  
DB 1141 GCGGAGCGATGATCATCAAGTGTCTCAACTGCGGCAAGAGGCGGCGGCGGCGGCG 1200  
QY 1201 TGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGGCGGCGGCGGCGG 1260  
DB 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGGCGGCGGCGG 1260  
QY 1258 GACTGACCGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGGCCAGCCCAAGAGGCG 1317  
DB 1258 GACTGACCGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGGCCAGCCCAAGAGGCG 1317  
QY 1321 CCGGCACTTCTGCAAGACCGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1377  
DB 1321 CCGGCACTTCTGCAAGACCGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1377  
QY 1378 CCGGCGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGGCGGCGGCGGCGGCGGCG 1437  
DB 1378 CCGGCGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGGCGGCGGCGGCGGCGGCG 1437  
QY 1438 GACCGGAGCCCTACCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1497  
DB 1438 GACCGGAGCCCTACCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1497  
QY 1498 CTGAGCCAGTAA 1509  
DB 1483 CTGAGCCAGTAA 1494  
  
RESULT 11  
ABL40020

ID ABL40020 standard; DNA; 1491 BP.  
XX  
AC ABL40020;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Synthetic Gag polynucleotide sequence SEQ ID NO:99.  
XX  
KW Human immunodeficiency virus type C; antigenic HIV type C protein;  
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
KW immunostimulant; gene therapy; gene; ds.  
XX  
OS Human immunodeficiency virus; type C.  
OS Synthetic.  
XX  
PN MO200204493-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
XX  
XX WPI; 2002-154920/20.  
DR  
XX  
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunization or generation of packaging cell  
PT lines, particularly in gene therapy.  
XX  
XX Claim 8; Fig 70; 233pp; English.  
PS  
XX  
XX The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (I). (I) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 1491 BP; 319 A; 557 C; 457 G; 158 T; 0 U; 0 Other;  
Query Match 85.8%; Score 1294.8; DB 6; Length 1491;  
Best Local Similarity 92.7%; Pred. No. 1e-150;  
Matches 1401; Conservative 0; Mismatches 87; Indels 24; Gaps 3;  
QY 1 ATGGGGCGCCCGCCGACATCCTGCGCGCGGAGAGCTGGACAAGTGGAGAGAATCCGC 60  
DB 1 ATGGGGCGCCCGCCGACATCCTGCGCGCGGAGAGCTGGACAAGTGGAGAGAATCCGC 60  
QY 61 CTGGCGCCCGCGGCAAGAGCACTATGCTGAAGCACTGTGTGTGGCCAGCCGCGAG 120  
DB 61 CTGGCGCCCGCGGCGGCAAGCACTATGCTGAAGCACTGTGTGTGGCCAGCCGCGAG 120  
QY 121 CTGGAAGGCTTCCGCTTGAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCATC 180  
DB 121 CTGGAAGGCTTCCGCGTGAACCCCGGCTGTGAGACCAAGCAAGGCTGCGCGCATC 180  
QY 181 ATGAAGCAGCTGACCGCCCGCTGCAAGCCGCAAGGAGAGTCCGCAAGCTTTCAAC 240  
DB 181 ATCAAGCAGCTGACCGCCCGCTGCAAGCCGCAAGGAGATCCGCAAGCTTTCAAC 240  
QY 241 ACCGTGGCCACCTTGTACTGCGTGAACGCGGCGCATCGAGTCCGCGACACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCTTGTACTGCGTGAACCAAGGCGCATCGTGGCGACCAAGAGAGGCC 300

QY 301 CTGGAAGAGATGAGAGAGAGAGCAACAAGTCCAGCAGAAAGCCAGAGGCCAGAGAG 360  
DB 301 CTGGAAGAGATGAGAGAGAGAGCAACAAGTCCAGCAGAAAGCCAGAGGCCAGAGAG 360  
QY 361 GCCGACGCGCAAGGTGAGCCAGAACTACCCCATCGTGAGAACTGTGAGGGCCAGATGGTG 420  
DB 361 GCCGACAGAAAGGTGAGCCAGAACTACCCCATCGTGAGAACTGTGAGGGCCAGATGGTG 420  
QY 421 CACGAGGCCATGACGCCCCCGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
DB 421 CACGAGGCCATGACGCCCCCGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
QY 481 TTCAAGCCCCGAGGTGATCCCAATGTTCAACCGCCCTGAGCGAGGGGCCACCCCGAGAG 540  
DB 481 TTCAAGCCCCGAGGTGATCCCAATGTTCAACCGCCCTGAGCGAGGGGCCACCCCGAGAG 540  
QY 541 CTGAACACGATGTTGAACACCGTGGGGCGGCCACAGGGCCGCAATGCATGCTGAAGAG 600  
DB 541 CTGAACACCATGCTGAACACCGTGGGGCGGCCACAGGGCCGCAATGCATGCTGAAGAG 600  
QY 601 ACCATCAACGAGAGAGCGCCCGCAAGTGGGACCGCCTGCACCCCGTGACGCGGCCCCGTG 660  
DB 601 ACCATCAACGAGAGAGCGCCCGCAAGTGGGACCGCCTGCACCCCGTGACGCGGCCCCGTG 660  
QY 661 GCGCGCGCGCAAGTGGCGGACCGCCCGCGGCAAGCATCGCGCGGCCACCAAGACCCCTG 720  
DB 661 GCGCGCGCGCAAGTGGCGGACCGCCCGCGGCAAGCATCGCGCGGCCACCAAGACCCCTG 720  
QY 721 CAGGAGCAGATCGCCTGGATGACCAACCAACCCCGCTGCGCGCGCGGCGCATCTTCAAG 780  
DB 721 CAGGAGCAGATCGCCTGGATGACCAACCAACCCCGCTGCGCGCGCGGCGCATCTTCAAG 780  
QY 781 CGGTGATCATCTTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCCGTGAGCATCTG 840  
DB 781 CGGTGATCATCTTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCCGTGAGCATCTG 840  
QY 841 GACATCCGCGAGGCGCCCAAGGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900  
DB 841 GACATCAAGCAGGAGCGCCCAAGGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900  
QY 901 CTGCGCGCGGAGAGAGCAACCAAGAGGTGAAGAACTGATGACCGACCCCTGTGCTG 960  
DB 901 CTGCGCGCGGAGAGAGCAACCAAGAGGTGAAGAACTGATGACCGACCCCTGTGCTG 960  
QY 961 CAGAACGCAACCCCGACTGCAAGACCATCTGCGGCTCTGCGCGCGCGCGGCGCACCTG 1020  
DB 961 CAGAACGCAACCCCGACTGCAAGACCATCTGCGGCTCTGCGCGCGCGCGGCGCACCTG 1020  
QY 1021 GAGGAGATGATGACCGCCTGCGAGGGCGTGGGGCGCCCGCGCAAGGCCGCGTGTG 1080  
DB 1021 GAGGAGATGATGACCGCCTGCGAGGGCGTGGGGCGCCCGCGCAAGGCCGCGTGTG 1080  
QY 1081 GCGGAGGCGATGAGCGAGGCAACAGCGTGAACATCATGATGAGAGAGCAACTTCAAG 1140  
DB 1081 GCGGAGGCGATGAGCGAGGCAACAGCGTGAACATCATGATGAGAGAGCAACTTCAAG 1140  
QY 1141 GCGCCCGCGGCGCAAGTCAAGTCTTCAACTGCGGCAAGGAGGCCACATCGCGCAAGAAC 1200  
DB 1141 GCGCCCGCGGCGCGTGAAGTCTTCAACTGCGGCGGCGGCGGAGGCCACATCGCGCAAC 1200  
QY 1201 TGCGCGCGCCCGCAAGAGGCTGTGAGTGCAGGAGAGGGCCACCAAGTGAAG 1260  
DB 1201 TGCGCGCGCCCGCAAGGCGGCTGTGAGTGCAGGAGAGGGCCACCAAGTGAAG 1260  
QY 1261 GACTGACCGAGCGCGCAAGCACTTCTGGGCAAGATCTGGCCAGCCACCAAGAGGCCG 1320  
DB 1261 GACTGACCGAGCGCGCAAGCACTTCTGGGCAAGATCTGGCCAGCCACCAAGAGGCCG 1320  
QY 1321 CCGGCAACTTCTTGAGAACCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
DB 1321 CCGGCAACTTCTTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380

QY 1378 CCCCCCGGAGAGCTTCCGCTTCGAGAGACCACCCCGCCCAAGCAGAGCCCAAG 1437  
 Db 1381 CCCCCCGGAGAGCTTCAAGTTCAGAGAGA-----CCCCCAAGCAGAGCCCAAG 1431  
 QY 1438 GACCGCGAGCCCTACCGCGAGCCCCCTGACCGCCCTGCGCAGCTGTTCCGCGAGCGGCCCC 1497  
 Db 1432 G-----ACCGCGAGCCCTGACCGCCTGAAGAGCCTGTTCCGCGAGCGGCCCC 1479  
 QY 1498 CTGAGCCAGTAA 1509  
 Db 1480 CTGAGCCAGTAA 1491  
 RESULT 12  
 ADM73833  
 ID ADM73833 standard; DNA; 1491 BP.  
 AC ADM73833;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE HIV-1 polynucleotide #68.  
 XX  
 KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
 KW HIV type C protein; immunostimulant.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 PN US2003223961-A1.  
 XX  
 PD 04-DEC-2003.  
 XX  
 PF 05-JUL-2001; 2001US-00899575.  
 XX  
 PR 05-JUL-2000; 2000US-00610313.  
 XX  
 PA (MEGE/) MEGEDE J Z.  
 PA (BARN/) BARNETT S W.  
 PA (ENGE/) ENGELBRECHT S.  
 PA (RENS/) RENSBURG E J V.  
 XX  
 PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;  
 XX  
 DR WPI; 2004-060515/06.  
 XX  
 PT New expression cassette comprising a polynucleotide sequence encoding an  
 PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
 PT immunization, generating of packaging cell lines or in producing HIV Type  
 PT C proteins.  
 XX  
 PS Claim 8; SEQ ID NO 99; 160pp; English.  
 XX  
 CC The invention relates to an expression cassette comprising a  
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
 CC also relates to a recombinant expression system for use in a host cell  
 CC comprising an expression cassette, where the polynucleotide sequence  
 CC further comprises control elements capable of driving expression in the  
 CC selected host cell, a cell comprising an expression cassette where the  
 CC polynucleotide sequence further comprises control elements compatible  
 CC with the expression in the cell and a composition for generating an  
 CC immunological response, comprising an expression cassette. The expression  
 CC cassette and the methods of the invention are useful in eliciting an  
 CC immune response, in DNA immunisation, in generation of packaging cell  
 CC lines and in producing HIV Type C proteins. This sequence represents an  
 CC HIV-1 polynucleotide of the invention.  
 XX  
 SQ Sequence 1491 BP; 319 A; 557 C; 457 G; 158 T; 0 U; 0 Other;  
 Query Match 85.8%; Score 1294.8; DB 12; Length 1491;  
 Best Local Similarity 92.7%; Pred. No. 1e-150;  
 Matches 1401; Conservative 0; Mismatches 87; Indels 24; Gaps 3;  
 QY 1 ATGGGCGCGCCGCGCCAGCATCTGCGCGCGGAGAGAGCTGACACAAGTGGAGAGATCCGC 60

Db 1 ATGGGCGCGCGCCGCGCCAGCATCTGCGCGCGGAGAGAGCTGACACAAGTGGAGAGATCCGC 60  
 QY 61 CTGCGCGCGCGCGCGCAAGACATAATGCTGAAGACACCTGGTGTGGGCCAGCCGCGAG 120  
 Db 61 CTGCGCGCGCGCGCGCGCAAGACATAATGCTGAAGACACCTGGTGTGGGCCAGCCGCGAG 120  
 QY 121 CTGGAAGGCTTTCGCTTGAACCCCGCGCTGCTGAAGACCCCGGAGGGCTGCAAGCAGATC 180  
 Db 121 CTGGAAGGCTTTCGCTTGAACCCCGCGCTGCTGAAGACCCCGGAGGGCTGCAAGCAGATC 180  
 QY 181 ATGAAGCAGCTGAGAGCCCGCGCTGCAAGACCCGCAACCGAGAGCTGCGCAGCTGTACAAC 240  
 Db 181 ATGAAGCAGCTGAGAGCCCGCGCTGCAAGACCCGCAACCGAGAGATCCGACGCTGTACAAC 240  
 QY 241 ACCGTGGCCACCTGTACTGCGGTGCAAGCGCCGATCGAGGTCCGCAACCAAGAGAGGCC 300  
 Db 241 ACCGTGGCCACCTGTACTGCGGTGCAAGCGCCGATCGAGGTCCGCAACCAAGAGAGGCC 300  
 QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCAGCAGAAACCAGCAGAGGCCCAAGAGAG 360  
 Db 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCAGCAGAAACCAGCAGAGGCCCAAGAGAG 360  
 QY 361 GCCGAGCGCAAGTGAAGCCAGAACTACCCCATCGTGCAGAACTTGACAGGGCCAGATGTTG 420  
 Db 361 GCCGAGCGCAAGTGAAGCCAGAACTACCCCATCGTGCAGAACTTGACAGGGCCAGATGTTG 420  
 QY 421 CACCAAGGCCATCAGCCCCCGGCAACCTGAACGCTGGGTGAAGGTATCGAGAGAGAGGCC 480  
 Db 421 CACCAAGGCCATCAGCCCCCGGCAACCTGAACGCTGGGTGAAGGTATCGAGAGAGAGGCC 480  
 QY 481 TTCAAGCCCCGAGGTGATCCCGCATGTTCAACCGCCCTGAGCGAGGGCGGCCACCCCCAGAGC 540  
 Db 481 TTCAAGCCCCGAGGTGATCCCGCATGTTCAACCGCCCTGAGCGAGGGCGGCCACCCCCAGAGC 540  
 QY 541 CTGAACACGATGTTGAACACCGGTGGCGGCCACCAAGCGGCCCATGACATGCTGAAGAGC 600  
 Db 541 CTGAACACGATGTTGAACACCGGTGGCGGCCACCAAGCGGCCCATGACATGCTGAAGAGC 600  
 QY 601 ACCATCAACGAGAGAGGCGCGGAGTGGGACCGGCTGCACCCCGTGCAGGCGCCCGCGTG 660  
 Db 601 ACCATCAACGAGAGAGGCGCGGAGTGGGACCGGCTGCACCCCGTGCAGGCGCCCGCGTG 660  
 QY 661 GCGCGCGCGGAGATGCGCGGACCCCGCGGCAAGCAATGCGCGCGGCCACCAAGCCTTG 720  
 Db 661 GCGCGCGCGGAGATGCGCGGACCCCGCGGCAAGCAATGCGCGCGGCCACCAAGCCTTG 720  
 QY 721 CAGGAGCAGATCGCCTGATGACCAAGCAACCCCGCATCCCGTGGCGGACATCTACAAG 780  
 Db 721 CAGGAGCAGATCGCCTGATGACCAAGCAACCCCGCATCCCGTGGCGGACATCTACAAG 780  
 QY 781 CGGTGATCATCTGCGGCTGAACAAGATCGTGGGATGTACAAGCCCGGTGAGCATCTG 840  
 Db 781 CGGTGATCATCTGCGGCTGAACAAGATCGTGGGATGTACAAGCCCGGTGAGCATCTG 840  
 QY 841 GACATCCGCGGAGGCGCCCAAGAGGCGCTTCCGCGACTACGTGAGCGGCTTCTTCAAGACC 900  
 Db 841 GACATCCGCGGAGGCGCCCAAGAGGCGCTTCCGCGACTACGTGAGCGGCTTCTTCAAGACC 900  
 QY 901 CTGCGCGCGGAGGAGGCGCAACCAAGAGCTGAAGAACTGATGACCGAGACCCCTGCTGGTG 960  
 Db 901 CTGCGCGCGGAGGAGGCGCAACCAAGAGCTGAAGAACTGATGACCGAGACCCCTGCTGGTG 960  
 QY 961 CAGAACGCCAACCCCGCACTGCAAGACCATCTGCGGCTCTCGGCGCCCGGCGCACCTTG 1020  
 Db 961 CAGAACGCCAACCCCGCACTGCAAGACCATCTGCGGCTCTCGGCGCCCGGCGCACCTTG 1020  
 QY 1021 GAGGAGATGATGACCGGCTGCAAGGCGGTGGGCGGCGCCGCAAGAGCGCGGCTG 1080  
 Db 1021 GAGGAGATGATGACCGGCTGCAAGGCGGTGGGCGGCGCCGCAAGAGCGCGGCTG 1080  
 QY 1081 GCCGAGCGGATGAGCGGAGGCGCAACGCTGAACATCATGATGACAGAGAGCAACTTCAAG 1140

Db 1081 GCCGAGCCATGAGCCAGGCCAACAACCAAGCGTGATGATCCAGAAGCAACTTCAAG 1140  
QY 1141 GGGCCCCGGCGCAACGTCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATCGCCAAAGAAC 1200  
Db 1141 GGGCCCCGGCGCGCGGTGAAGTGTCTTCAACTGCGGCGCGGAGGGCCACATCGCCCCGAAC 1200  
QY 1201 TGCCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCATGATGAAG 1260  
Db 1201 TGCCCGCGCCCCCGCAAGCGCGGTGCTGGAAGTGGCGCAAGAGGGCCACCATGATGAAG 1260  
QY 1261 GACTGCAACCGAGCGCGCAAGCCCACTTCTGGGCAAGATCTGGCCAGCCACAAGGGCGGC 1320  
Db 1261 GACTGCAACCGAGCGCGCAAGCCCACTTCTGGGCAAGATCTGGCCAGCCACAAGGGCGGC 1320  
QY 1321 CCGGCAACTTCTGCAAGAACCGCAGCGAGCCCGCGCCGCCACC--GTGCCACCGGC 1377  
Db 1321 CCGGCAACTTCTGCAAGAGCGCGCCGAGCCACCGCCCCCTGGAGCCCAACCGGC 1380  
QY 1378 CCGCGCGCGAGAGCTTCCGCTTCCAGAGAACCAACCCCGCCCCCAAGCAGAGCCCAAG 1437  
Db 1381 CCGCGCGCGAGAGCTTCAAGTTCAAGAGA-----CCCCCAAGCAGAGCCCAAG 1431  
QY 1438 GACCGCGAGCCCTTACCGCGAGCCCTGACCGCCCTGCGAGCCTGTTGGCAGCGCGCC 1497  
Db 1432 G-----ACCGCGAGCCCTGACCGCTGAAGAGCCTGTTGGCAGCGAGCCCC 1479  
QY 1498 CTGAGCCAGTAA 1509  
Db 1480 CTGAGCCAGTAA 1491

RESULT 13  
AAAS1609  
ID AAAS1609 standard; DNA; 1479 BP.  
XX AAAS1609;  
AC

DT 31-OCT-2000 (first entry)  
XX  
DE HIV synthetic Gag polynucleotide.

XX Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;  
KM DNA immunization; packaging cell line; antigen presentation; ss.

XX Human immunodeficiency virus; type C strain AF110965.  
OS Synthetic.

XX Key Location/Qualifiers  
FH CDS 1..1479  
FT /\*tag= a  
FT /product= "Synthetic Gag"  
FT /note= "Codon usage pattern was modified and inhibitory  
FT elements (INS) and RRE sites were inactivated resulting  
FT in improved expression"  
XX

PN WO200039304-A2.

XX 06-JUL-2000.

PD 30-DEC-1999; 99WO-US031273.

XX 31-DEC-1998; 98US-0114495P.

PR 01-SEP-1999; 99US-0152195P.

XX (CHIR ) CHIRON CORP.

XX Barnett S, Zur Megede J;

DR WPI; 2000-452401/39.

DR P-PSDB; AAY96943.

XX Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env  
PT polypeptide and the polypeptide useful for immunizing a mammal especially

PT human against HIV.  
XX  
XX Claim 2; Page 92-93; 113pp; English.  
PS  
XX Expression cassettes comprising a polynucleotide encoding antigenic type  
CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
CC in DNA immunization, generation of packaging cell lines and production of  
CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
CC cassettes exhibit increased potency for induction of cytotoxic T-  
CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-  
CC assemble into non-infectious virus-like particles which are used as a  
CC matrix for the proper presentation of an antigen entrapped or associated  
CC to the immune system of the host  
XX

SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;

Query Match 84.6%; Score 1276.8; DB 3; Length 1479;  
Best Local Similarity 92.5%; Pred. No. 1.7e-148;  
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;

QY 1 ATGGCGCGCCGCGCAGCATCTCGCGCGCGGAGAGCTGGAACAAGTGGAGAGATCCGC 60  
Db 1 ATGGCGCGCCGCGCAGCATCTCGCGCGCGGAGAGCTGGAACGCTGGAGCGCATCCGC 60  
QY 61 CTGCGCGCGCGGCAAGAGCACTACATGCTGAAGCACCTGTGTGGCCAGCCGAG 120  
Db 61 CTGCGCGCGCGGCAAGAGTGTACATGATGAAGCACCTGTGTGGCCAGCCGAG 120  
QY 121 CTGAGAGGCTTGCCTTGAACCCCGCCTGCTGAGAACCGCGAGGCTGCAAGCAGATC 180  
Db 121 CTGAGAGAGTTGCTGCTTGAACCCCGCCTGCTGAGAACCGAGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGAGCGCCGCTGAGACCGGCAACGAGAGCTGCGACCTGTACAAC 240  
Db 181 ATCCGCAAGCTGACACCCGCTGAGACCGGCAACGAGAGCTGAAGCCTGTACAAC 240  
QY 241 ACCGTGGCACCCTGTACTGCTGACACCGGCGCATCGAGTCCGCGACCAAGAGGCC 300  
Db 241 ACCGTGGCACCCTGTACTGCTGACACCGGCGCATCGAGTCCGCGACCAAGAGGCC 300  
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCAGCAGAAAGCCAGAGCCAGAG 360  
Db 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCAGCAGAAAGTCCAGAGCCAGAGCC 360  
QY 361 GCCGAC--GGCAAGGTGAGCCGAACCTACCCCATGTCAGAACTTGAGGGCCAGATG 417  
Db 361 GCCGACAGAGGCAAGGTGAGCCGAACCTACCCCATGTCAGAACTTGAGGGCCAGATG 420  
QY 418 GTGACCAAGGCGCATCAGCCCGGACCTGAAGCCTGGTGAAGTGAAGTGAAG 477  
Db 421 GTGACCAAGGCGCATCAGCCCGGACCTGAAGCCTGGTGAAGTGAAGTGAAG 480  
QY 478 GCCTTACGCGCGGAGTATCCCATGTTCACCGCCCTGAGCGAGGGGCCACCCCCAG 537  
Db 481 GCCTTACGCGCGGAGTATCCCATGTTCACCGCCCTGAGCGAGGGGCCACCCCCAG 540  
QY 538 GACCTGAACAAGATGTTGAACACCGTGGCGGCCACAGGCGCCATGCAAGTGTGAAG 597  
Db 541 GACCTGAACAAGATGTTGAACACCGTGGCGGCCACAGGCGCCATGCAAGTGTGAAG 600  
QY 598 GACACCATCAACGAGAGGCGCGGAGTGGAGCCGCTGCAACCCCGTGAGGCGCGCC 657  
Db 601 GACACCATCAACGAGAGGCGCGGAGTGGAGCCGCTGCAACCCCGTGAGGCGCGCC 660  
QY 658 GTGGCGCGCGCGAGATGCGCGACCCCGCGGAGCGACATCGCGCGGCCACAGCACC 717  
Db 661 ATGCGCGCGCGCGAGATGCGCGACCCCGCGGAGCGACATCGCGCGGCCACAGCACC 720  
QY 718 CTGCAAGAGAGATGCGCTGATGAACCAAGACCCCGCGTGCCTGGCGGACATCTAC 777  
Db 721 CTGCAAGAGAGATGCGCTGATGAACCAAGACCCCGCGTGCCTGGCGGACATCTAC 780  
QY 778 AAGCGGTGATCATCTCGGCTGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC 837

```
Db 781 AAGCGTGATCATCTGGGCTGAGACAGATCGTGGGATGTACAGCCCGGTGAGCATC 840
QY 838 CTGGACATCCGCGAGGGCCCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTTAAG 897
Db 841 CTGGACATCAAGCAGGGCCCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTTAAG 900
QY 898 ACCCTGCGCGCGAGCAGGCGCAACCCAGACGTGAAGACTGTGATGACCCGACCTGCTG 957
Db 901 ACCCTGCGCGCGAGCAGGCGCAACCCAGAGGTGAAGACTGTGATGACCCGACCTGCTG 960
QY 958 GTGCAGAACGCCCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCGCAC 1017
Db 961 GTGCAGAACGCCCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCGCAC 1020
QY 1018 CTGGAGAGATGATGACCCGCTGCGAGGCGTGGCGGCCCGCGCACAGGCCCGCGCTG 1077
Db 1021 CTGGAGAGATGATGACCCGCTGCGAGGCGTGGCGGCCCGCGCACAGGCCCGCGCTG 1080
QY 1078 CTGGCCGAGCGATGAGCGCAAGCCCAAGCGTGAACATCATGATGACAGAGCACTTC 1137
Db 1081 CTGGCCGAGCGATGAGCGCAAGCCCAAGCGTGAACATCATGATGACAGAGCACTTC 1137
QY 1138 AAGGGCCCCCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGGGGCCACATCGCCCAAG 1197
Db 1138 AAGGGCCCCCGCGCATGCTCAAGTCTTCAACTGCGGCAAGAGGGGCCACATCGCCCGC 1197
QY 1198 AACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGGGCCACCAAGATG 1257
Db 1198 AACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGGGCCACCAAGATG 1257
QY 1258 AAGGACTGCAACCGAGCGCGCAACTTCTGCGCAAGATCTGCGCCAGCCCAAGAGGGC 1317
Db 1258 AAGGACTGCAACCGAGCGCGCAACTTCTGCGCAAGATCTGCGCCAGCCCAAGAGGGC 1317
QY 1318 CGCCCCGCGCAACTTCTGCGCAAGACCGGAGCGAGCCCGCGCCCAACCGTGGCCCAAGCC 1377
Db 1318 CGCCCCGCGCAACTTCTGCGAGAGCGG-----CGCCGAGCCCAAGCCCGC 1359
QY 1378 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCCCAAGCAGAGGCCCAAG 1437
Db 1360 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCCCAAGCAGAGGCCAAG 1419
QY 1438 GACCGGAGCGCTTACCGCGAGCGCCCTGACCGCCCTGCGCAGCTGTTGGGAGCGGCC 1497
Db 1420 G-----ACCGCGAGACCTTGAACCAAGCTGTAAGAGCTGTTGGGCAAGCAAGCC 1467
QY 1498 CTGAGCCAGTAA 1509
Db 1468 CTGAGCCAGTAA 1479

RESULT 14
ID AAL44548 standard; DNA; 1479 BP.
XX AAL44548;
AC
XX
DT 29-AUG-2003 (revised)
DT 08-NOV-2002 (first entry)
XX
DE HIV-1 p55gag polypeptide coding sequence 1.
XX
KW HIV; ds; vaccine; gene; immune response; microparticle;
KW adsorbent surface; poly(alpha-hydroxy acid); polyhydroxy butyric acid;
KW polycaprolactone; polyorthoester; polycyanacrylate; detergent;
KW submicron emulsion; viral infection; bacterial infection;
KW parasitic infection; HIV-1 p55gag polypeptide.
OS
XX Human immunodeficiency virus 1.
PN
XX WO200226209-A2.
```

```
PD 04-APR-2002.
XX
XX 28-SEP-2001; 2001WO-US030540.
XX
XX 28-SEP-2000; 2000US-0236105P.
PR 30-AUG-2001; 2001US-0315905P.
XX
XX (CHIR ) CHIRON CORP.
XX
PI O'hagan D, Otten G, Donnelly J, Polo JM, Barnett S, Singh M,
PI Ulmer J, Dubensky TW;
XX
XX WPI; 2002-519084/55.
XX
PT A microparticle to which a biologically active macromolecule is adsorbed,
PT for use as a vaccine composition to treat viral, bacterial or parasitic
PT infections, comprises a polymer microparticle, a detergent and a
PT submicron emulsion.
XX
XX Claim 72; Fig 1; 100pp; English.
XX
XX The invention relates to a method of raising an immune response in a host
XX animal. The method of the invention comprises administering a
XX microparticle that has an adsorbent surface to which a first biologically
XX active macromolecule (e.g. a nucleic acid) has been adsorbed. The
XX microparticle comprises a polymer microparticle of poly(alpha-hydroxy
XX acid), a polyhydroxy butyric acid, a polycaprolactone, a polyorthoester,
XX a polycyanacrylate, a detergent, and submicron emulsion. The method/
XX microparticle of the invention is useful for immunising a host animal
XX against viral, bacterial or parasitic infections. The present DNA
XX sequence encodes a HIV-1 p55gag polypeptide. (Updated on 29-AUG-2003 to
XX standardise OS field)
SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;
```

```
Query Match 84.6%; Score 1276.8; DB 6; Length 1479;
Best Local Similarity 92.5%; Pred. No. 1.7e-148;
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;
```

```
QY 1 ATGGGCGCCCGCGCGAGCATCTGCGCGGCGAGAGCTGGAACAAGTGGAGAAAGATCCGC 60
Db 1 ATGGGCGCCCGCGCGAGCATCTGCGCGGCGAGAGCTGGAACGCTGGAGCGCATCCGC 60
QY 61 CTGCGCCCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCCAGCCGCGAG 120
Db 61 CTGCGCCCGCGCGCGCAAGAGTGTCTACATGATGAAGCACTGTGTGTGGGCCAGCCGCGAG 120
QY 121 CTGGAGGGCTTCCGCTTGAACCCCGGCTGTGAGACCGCGGAGGGCTGCAAGCAGATC 180
Db 121 CTGGAGAGTTCGCGCTTGAACCCCGGCTGTGAGACCGCGGAGGGCTGCAAGCAGATC 180
QY 181 ATGAAGCAGCTGACAGCCCGCTGCAAGACCGGACCGAGGAGCTGCGCAGCTGTACAA 240
Db 181 ATCCGCCAGCTGACACCCCGCTGCAAGACCGGACCGAGGAGAGCTGAAGAGCTGTTCAC 240
QY 241 ACCGTGGCCACCTGTACTGCTGTCAGCGCGGATGAGGTCCGCGACCAAGAGAGGCC 300
Db 241 ACCGTGGCCACCTGTACTGCTGTCAGAGAGAGATGAGGTCCGCGACCAAGAGAGGCC 300
QY 301 CTGACACAAGATGAGAGAGAGAGACAAGTCCCGACAGAGAACCCAGAGGCCAAGAG 360
Db 301 CTGACACAAGATGAGAGAGAGAGACAAGTCCCGACAGAGAGATCCAGAGAGGCCAGGCC 360
QY 361 GCCGAC---GGCAAGTGAAGCCAGAACTACCCCATGTGCAAGAACTGCAAGGGCCAGATG 417
Db 361 GCCGACAAAGGGCAAGGTGAAGCCAGAACTACCCCATGTGCAAGAACTGCAAGGGCCAGATG 420
QY 418 GTGACACAGGCCATACAGCCCGCGACCTGAACGCTGGGTGAAGGTGATCGAGGAGAG 477
Db 421 GTGACACAGGCCATACAGCCCGCGACCTGAACGCTGGGTGAAGGTGATCGAGGAGAG 480
QY 478 GCCTTCAGCCCGGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCAACCCCGCAG 537
Db 478 GCCTTCAGCCCGGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCAACCCCGCAG 537
```

```
Db 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCCCAACCCCCAG 540
QY 538 GACCTGAACACGATGTTGAACACCGTGGGCGGCCAACAGGCCCGCATGAGATGCTGAAG 597
Db 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCCAACAGGCCCGCATGAGATGCTGAAG 600
QY 598 GACACCATCAACGAGAGAGGCCCGCCGAGTGGGACCGCCTTGCAACCCCGTGAGGCCGCGCC 657
Db 601 GACACCATCAACGAGAGAGGCCCGCCGAGTGGGACCGCCTTGCAACCCCGTGAGGCCGCGCC 660
QY 658 GTGGCCCCCGCCGAGATGCGCGACCCCGCGGCGAGCGACATGCGCGCGCCACCAAGACC 717
Db 661 ATGCCCCCGCCGAGATGCGCGAGCCCGCGGCGAGCATGCGCGGCAACCAAGCAACC 720
QY 718 CTGACGAGCAGATCGCCTGGATGACCAACCCCGCGTGCCCGTGCGGCGACATCTAC 777
Db 721 CTGACGAGCAGATCGCCTGGATGACCAACCCCGCATCCCGTGCGGCGACATCTAC 780
QY 778 AAGCGGTGATCATCTTGGGCTGAAACAAGATCTGCGGATGTACAAGCCCGTGAGCATC 837
Db 781 AAGCGGTGATCATCTTGGGCTGAAACAAGATCTGCGGATGTACAAGCCCGTGAGCATC 840
QY 838 CTGACATCCGCCAGGGCCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 897
Db 841 CTGACATCAAGCAGGGCCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 900
QY 898 ACCCTGCGCGCCGAGCGACGCCAACCCAGAGCTGAAGAACTGATGACCGAGACCCCTG 957
Db 901 ACCCTGCGCGCCGAGCGAGACGCCAACCCAGAGCTGAAGAACTGATGACCGAGACCCCTG 960
QY 958 GTGCAAGAACCCAAACCCGACTGCAAGAACCATCTGCGCGCTCTGCGCCCGCGCCACC 1017
Db 961 GTGCAAGAACCCAAACCCGACTGCAAGAACCATCTGCGCGCTCTGCGCCCGCGCCACC 1020
QY 1018 CTGAGAGAGATGATGACCGCCTGCGCAAGGCGGTGCGGCCCCGCGCAAGGCCCGGTG 1077
Db 1021 CTGAGAGAGATGATGACCGCCTGCGCAAGGCGGTGCGGCCCCGCGCAAGGCCCGGTG 1080
QY 1078 CTGCGCGAGGCGATGAGCCAGGCCAACAGCGTGAAACATCATGATGACAGAAAGCACTTC 1137
Db 1081 CTGCGCGAGGCGATGAGCCAGGCCAACACC--AGCGTATGATGACAGAAAGCACTTC 1137
QY 1138 AAGGCCCCCGCGCAAGTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCAG 1197
Db 1138 AAGGCCCCCGCGCGCATGTCAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCCGC 1197
QY 1198 AACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAGTGCGGCAAGAGGGCCACCATG 1257
Db 1198 AACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAGTGCGGCAAGAGGGCCACCATG 1257
QY 1258 AAGGACTGACCGAGCGCCAGCCAACTTCTGGGCAAGATTGCGCCACAGCCAGGGGC 1317
Db 1258 AAGGACTGACCGAGCGCCAGCCAACTTCTGGGCAAGATTGCGCCACAGCCAGGGGC 1317
QY 1318 CGCCCCGGCACTTCTGTCAGAACCGGAGCGCGCCCGCCCAACCGTGCCCAACGCC 1377
Db 1318 CGCCCCGGCACTTCTGTCAGAGCGG-----CGCCGAGCCCAACGCC 1359
QY 1378 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCCCAAGAGAGCCCAAG 1437
Db 1360 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCAAGAGAGAGCAAG 1419
QY 1438 GACCGGAGAGCCCTACCGCGAGCCCGCTGACCGCCCTGCGCAGCCTGTTCGCGAGCGGCC 1497
Db 1420 G-----ACCGGAGAGCCCTTGAACCAAGCCTGAAGAGCCTGTTCGCGCAACGACCCC 1467
QY 1498 CTGAGCCAGTAA 1509
Db 1468 CTGAGCCAGTAA 1479
```

RESULT 15  
ABL39953

```
ID ABL39953 standard; DNA; 1479 BP.
XX
AC ABL39953;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic Gag polynucleotide sequence SEQ ID NO:3.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR ) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX WPI; 2002-154920/20.
XX
DR New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
XX Example 1; Fig 1; 233pp; English.
XX
XX The present invention describes expression cassettes comprising a
XX polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
XX type C polypeptides. The expression cassettes comprise any of the HIV
XX type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
XX (I). (I) have immunostimulant activity and can be used in gene therapy.
XX The HIV type C polynucleotides are useful in applications including DNA
XX immunisation, generation of packaging cell lines, and production of HIV
XX type C proteins. The polynucleotides are particularly useful in gene
XX therapy and DNA immunisation applications. ABL39942 to ABL40054 and
XX ABB06204 to ABB06215 represent sequences used in the exemplification of
XX the present invention
SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;
Query Match 84.6%; Score 1276.8; DB 6; Length 1479;
Best Local Similarity 92.5%; Pred. No. 1.7e-148;
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;
QY 1 ATGGGCGCGCCCGCCAGCATCTCTGCGCGCGCGGAGAACTGGAACAAGTGGGAGAATCCGC 60
Db 1 ATGGGCGCGCCCGCCAGCATCTCTGCGCGCGCGGAGAACTGGAACAAGTGGGAGAATCCGC 60
QY 61 CTGCGCCCCCGCGCAAGAGCACTACATGCTGAAGCAACCTGTGTGGCCAGCCCGGAG 120
Db 61 CTGCGCCCCCGCGCAAGAGTGTACTATGATGAAGCAACCTGTGTGGCCAGCCCGGAG 120
QY 121 CTGAGGGGCTTCCGCTTGAACCCCGCGCTGTCGAGAACCGCCGAGGGCTGCAAGCATC 180
Db 121 CTGAGGAAGTTCCGCTTGAACCCCGCGCTGTCGAGAACCGCCGAGGGCTGCAAGCATC 180
QY 181 ATGAAGCAGCTGACAGCCCGCCCTGCAAGACCGGCAACGAGAGAGCTGCGCAGCCTGTAAAC 240
Db 181 ATCCGCCAGCTGACAGCCCGCCCTGCAAGACCGGCAACGAGAGAGCTGTAAGCCTGTAAAC 240
QY 241 ACCGTGGCCACCTGTACTGCGTGTGACCGCGGCAATCGAGGTCCGCGACCAAGAGAGGCC 300
Db 241 ACCGTGGCCACCTGTACTGCGTGTGACGAGAAAGATCGAGGTCCGCGACCAAGAGAGGCC 300
```

QY 301 CTGACAAGATCGAGAGGACAGACAACTCCACGAGAGAACCCAGCGCCCAAGGAG 360  
 Db 301 CTGACAAGATCGAGAGGACAGACAACTCCACGAGAGATCCAGCGAGCCGAGGCC 360  
 QY 361 GCGGAC---GGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTTGAGGGCCAGATG 417  
 Db 361 GCGGACAAAGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTTGAGGGCCAGATG 420  
 QY 418 GTGCACCAAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTATCGAGAGAG 477  
 Db 421 GTGCACCAAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTATCGAGAGAG 480  
 QY 478 GCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGGCCACCCCCAG 537  
 Db 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGGCCACCCCCAG 540  
 QY 538 GACCTGAACACGATGTTGAACACGTTGGGGGCCACCAAGGCCCGCATGAGATGCTGAAG 597  
 Db 541 GACCTGAACACGATGTTGAACACGTTGGGGGCCACCAAGGCCCGCATGAGATGCTGAAG 600  
 QY 598 GACACCATCAACGAGAGAGCCCGCGAGTGGGACCGCTGCACCCCGTGAGAGCGCGGCC 657  
 Db 601 GACACCATCAACGAGAGAGCCCGCGAGTGGGACCGCTGCACCCCGTGAGAGCGCGGCC 660  
 QY 658 GTGGCCCCCGGCGAGATGCGCGAGACCCCGCGGCGAGCGACATCGCCGCGGCCACCAAG 717  
 Db 661 ATGCCCCCGGCGAGATGCGCGAGACCCCGCGGCGAGCGACATCGCCGCGGCCACCAAG 720  
 QY 718 CTGACAGACAGATCGCTGATGATGACAGCAACCCCGCTGCCCCGTGGGGGACATCTAC 777  
 Db 721 CTGACAGACAGATCGCTGATGATGACAGCAACCCCGCTGCCCCGTGGGGGACATCTAC 780  
 QY 778 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 837  
 Db 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 840  
 QY 838 CTGACATCCGCGAGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTCAAG 897  
 Db 841 CTGACATCAAGCAGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTCAAG 900  
 QY 898 ACCCTGCGCGCGAGCGACGCCACCAAGGAGCTGAAGAACTGATGACCGAGACCCCTGCTG 957  
 Db 901 ACCCTGCGCGCGAGCGAGCAGCAACCAAGGAGTGAAGAACTGATGACCGAGACCCCTGCTG 960  
 QY 958 GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCCCCCGCGCGCAC 1017  
 Db 961 GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCCCCCGCGCGCAC 1020  
 QY 1018 CTGAGAGAGATGATGACCGCTGCGAGGGCGTGGGGCCCCCGGCCAACAGGCCCGCTG 1077  
 Db 1021 CTGAGAGAGATGATGACCGCTGCGAGGGCGTGGGGCCCCCGGCCAACAGGCCCGCTG 1080  
 QY 1078 CTGCGCGAGGCGATGAGCCAGGCGCAACGCTGAACATCATGATGCAAGAAGCACTTC 1137  
 Db 1081 CTGCGCGAGGCGATGAGCCAGGCGCAAC--AGCGTATGATGCAAGAAGCACTTC 1137  
 QY 1138 AAGGGCCCCCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCAAG 1197  
 Db 1138 AAGGGCCCCCGCGCATCTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGC 1197  
 QY 1198 AACTGCGCGCCCCCGCAAGAGGGTCTGTGAAGTGCGGCAAGAGGGCCACAGATG 1257  
 Db 1198 AACTGCGCGCCCCCGCAAGAGGGTCTGTGAAGTGCGGCAAGAGGGCCACAGATG 1257  
 QY 1258 AAGGACTGACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCTGGCCAGCGCAAGAGGC 1317  
 Db 1258 AAGGACTGACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCTGGCCAGCGCAAGAGGC 1317  
 QY 1318 CGCCCCGGCAACTTCTGCAAGACCGAGCGAGCCCGCCCGCACCGTGGCCACCGCC 1377  
 Db 1318 CGCCCCGGCAACTTCTGCAAGACCG-----CGCCGAGCCACCGCC 1359

QY 1378 CCCCCCGGAGAGCTTCCGCTTGAGAGAGACCACCCCCGCCCCCAAGAGAGCCCAAG 1437  
 Db 1360 CCCCCCGGAGAGCTTCCGCTTGAGAGAGACCACCCCCGGCCAGAGAGAGCAAG 1419  
 QY 1438 GACCGGAGCCCTACCGCGAGCCCCCTGACCGCCCTGCGAGGCTGTTGCGAGCGCCCC 1497  
 Db 1420 G-----ACCGGAGACCTTGACCAAGGCTGTTGCGCAAGAGCCCC 1467  
 QY 1498 CTGAGCCAGTAA 1509  
 Db 1468 CTGAGCCAGTAA 1479

Search completed: March 12, 2005, 17:35:57  
 Job time : 862.019 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 13:16:01 ; Search time 6734.44 Seconds  
(without alignments)  
10857.457 Million cell updates/sec

Title: US-09-475-704A-4

Perfect score: 1509  
Sequence: 1 atggcgcccgccgcccagcat.....gcgccccctgagccagtaa 1509

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1509	100.0	1509	6 AX455888	AX455888 Sequence
2	1507.4	99.9	1509	6 AX468544	AX468544 Sequence
3	1504.2	99.7	1509	6 AX468548	AX468548 Sequence
4	1489.8	98.7	1509	6 AX455905	AX455905 Sequence
5	1297.6	86.0	1494	6 AX455935	AX455935 Sequence
6	1294.8	85.8	1491	6 AX455983	AX455983 Sequence
7	1291.4	85.6	1485	12 AY181195	AY181195 Synthetic
8	1276.8	84.6	1479	6 AX455887	AX455887 Sequence
9	1276.8	84.6	1479	6 AX468543	AX468543 Sequence
10	1271.2	84.2	1479	6 AX468547	AX468547 Sequence
11	1260.8	83.6	1479	6 AX455904	AX455904 Sequence
12	1202.2	79.7	4288	6 AX149648	AX149648 Sequence
13	1201	79.6	1509	12 AF201927	AF201927 Synthetic
14	1201	79.6	1515	6 BD263637	BD263637 Improved
15	1201	79.6	1515	6 CQ870496	CQ870496 Sequence
16	1201	79.6	1515	6 AR373320	AR373320 Sequence
17	1201	79.6	4472	6 BD263697	BD263697 Improved
18	1201	79.6	4472	6 CQ870567	CQ870567 Sequence
19	1201	79.6	4472	6 AR373380	AR373380 Sequence

20	1201	79.6	4608	6 BD263698	BD263698 Improved
21	1201	79.6	4608	6 CQ870568	CQ870568 Sequence
22	1201	79.6	4608	6 AR373381	AR373381 Sequence
23	1201	79.6	4689	6 BD263696	BD263696 Improved
24	1201	79.6	4689	6 CQ870566	CQ870566 Sequence
25	1201	79.6	4766	6 AR373379	AR373379 Sequence
26	1201	79.6	4766	6 BD263695	BD263695 Improved
27	1201	79.6	4766	6 CQ870565	CQ870565 Sequence
28	1201	79.6	4766	6 AR373378	AR373378 Sequence
29	1199	79.5	2031	6 BD263640	BD263640 Improved
30	1199	79.5	2031	6 CQ870499	CQ870499 Sequence
31	1199	79.5	2031	6 AR373323	AR373323 Sequence
32	1195.8	79.2	1503	12 AY531263	AY531263 Synthetic
33	1186	78.6	9166	6 AX427930	AX427930 Sequence
34	1170.8	77.6	9788	6 AX427936	AX427936 Sequence
35	1145	75.9	1539	6 AX306431	AX306431 Sequence
36	1143.2	75.8	1548	12 AF287354	AF287354 Synthetic
37	1119	74.2	6438	6 AX427919	AX427919 Sequence
38	1119	74.2	9194	6 AX427925	AX427925 Sequence
39	1119	74.2	9194	6 AX427926	AX427926 Sequence
40	1119	74.2	12411	6 AX427927	AX427927 Sequence
41	1117.8	74.1	1847	12 AF202464	AF202464 Synthetic
42	1117.8	74.1	1847	12 AF202465	AF202465 Synthetic
43	1117.8	74.1	1853	6 BD263638	BD263638 Improved
44	1117.8	74.1	1853	6 CQ870497	CQ870497 Sequence
45	1117.8	74.1	1853	6 AR373321	AR373321 Sequence

ALIGNMENTS

RESULT 1	AX455888	1509 bp	DNA	linear	PAT 06-JUL-2002
LOCUS	AX455888	Sequence 4 from Patent WO0204493.			
DEFINITION	AX455888				
ACCESSION	AX455888				
VERSION	AX455888.1	GI:21714881			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE	1				
AUTHORS	zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.				
TITLE	Polynucleotides encoding antigenic hiv type c polypeptides,				
JOURNAL	Patent: WO 0204493-A 4 17-JAN-2002;				
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)					
FEATURES					
SOURCE	1..1509				
	/organism="synthetic construct"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:32630"				
	/note="synthetic Gag of HIV strain AF110967"				
ORIGIN					
Query Match	100.0%;	Score 1509;	DB 6;	Length 1509;	
Best Local Similarity	100.0%;	Pred. No. 3.2e-150;			
Matches 1509;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	ATGGCGCCCGCGCGCATCTCTGGCGCGGAGAGCTGGACAAGTGGAGAGATCCGC	60		
Db	1	ATGGCGCCCGCGCGCATCTCTGGCGCGGAGAGCTGGACAAGTGGAGATCCGC	60		
QY	61	CTGGCGCCCGCGCGCAAGACACTACATGCTGAAGCACCTGTGTGGCCAGCCGCGAG	120		
Db	61	CTGGCGCCCGCGCGCAAGACACTACATGCTGAAGCACCTGTGTGGCCAGCCGCGAG	120		
QY	121	CTGAGGGCTTCGCTTGAACCCGCGCTGCTGAGACCGCCGAGGGCTGCAAGCAGATC	180		
Db	121	CTGAGGGCTTCGCTTGAACCCGCGCTGCTGAGACCGCCGAGGGCTGCAAGCAGATC	180		
QY	181	ATGAAGCAGCTGAGCCCGCCCTGCGAGACCGGCAACGAGAGCTGCGAGCTGTACAC	240		

Db 181 ATGAAGCAGCTGACGCCCCCTGACAGCCGGCACCGAGGAGCTGCGACCTGTACAAC 240  
QY 241 ACCGTGGCCACCCTGTACTGCGTGACAGCCGGCATGAGGTCCGCGACACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCCTGTACTGCGTGACAGCCGGCATGAGGTCCGCGACACCAAGAGGCC 300  
QY 301 CTGGAACAAGATCGAGGAGAGCAGAAACAAGTCCAGCAGAGAACCAGCAGGCCCAAGAG 360  
Db 301 CTGGAACAAGATCGAGGAGAGCAGAAACAAGTCCAGCAGAGAACCAGCAGGCCCAAGAG 360  
QY 361 GCCGAGCGCAAGGTGAGCCAGAACTACCTCATCGTGCAAGAACTTGACAGGGCCAGATG 420  
Db 361 GCCGAGCGCAAGGTGAGCCAGAACTACCTCATCGTGCAAGAACTTGACAGGGCCAGATG 420  
QY 421 CACCAAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGGAGAGGCC 480  
Db 421 CACCAAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGGAGAGGCC 480  
QY 481 TTCAAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGGCCACCCCCAGGAC 540  
Db 481 TTCAAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGGCCACCCCCAGGAC 540  
QY 541 CTGAACAACGATGTTGAACACCGTGGGGGGCCACACGAGCCGCCATGAGATGCTGAAGGAC 600  
Db 541 CTGAACAACGATGTTGAACACCGTGGGGGGCCACACGAGCCGCCATGAGATGCTGAAGGAC 600  
QY 601 ACCATCAACGAGGAGGCGCCGAGTGGGACCGCCCTGCAACCCCGTGACAGGCCCGCCGTG 660  
Db 601 ACCATCAACGAGGAGGCGCCGAGTGGGACCGCCCTGCAACCCCGTGACAGGCCCGCCGTG 660  
QY 661 GCCCGCGGCCAGATGCGCGGACCCCGCGGCGAGCGACATCGCCGCGGCCACACGACCCCTG 720  
Db 661 GCCCGCGGCCAGATGCGCGGACCCCGCGGCGAGCGACATCGCCGCGGCCACACGACCCCTG 720  
QY 721 CAGGAGCAGATCGCTGGATGACCAAGCAACCCCGCTGCGCGTGGGGGACATCTACAAG 780  
Db 721 CAGGAGCAGATCGCTGGATGACCAAGCAACCCCGCTGCGCGTGGGGGACATCTACAAG 780  
QY 781 CGGTGGATCATCTGGGGCTTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTCTG 840  
Db 781 CGGTGGATCATCTGGGGCTTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTCTG 840  
QY 841 GACATCCGCCAAGGCGCCCAAGGAGGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900  
Db 841 GACATCCGCCAAGGCGCCCAAGGAGGCGCTTCCGCGACTACGTGACCGCTTCTTCAAGACC 900  
QY 901 CTGGCGCGCCGAGGAGGCGCCACCGCAGACGTGAAGAACTGATGACCGAGACCCCTGTG 960  
Db 901 CTGGCGCGCCGAGGAGGCGCCACCGCAGACGTGAAGAACTGATGACCGAGACCCCTGTG 960  
QY 961 CAGAAACGCCAACCCCGACTGCAAGAACCATCTGCGCGCTCTCGGCCCCGCGCGCACCCCTG 1020  
Db 961 CAGAAACGCCAACCCCGACTGCAAGAACCATCTGCGCGCTCTCGGCCCCGCGCGCACCCCTG 1020  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGGGCGTGGGCGGCCCCGCGCACAGGCGCGCTGCTG 1080  
Db 1021 GAGGAGATGATGACCGCTGCGCAGGGGCGTGGGCGGCCCCGCGCACAGGCGCGCTGCTG 1080  
QY 1081 GCCGAGGCGATGAGCCAGGCGCAACAGCGTGAACATCATGATGCAAGAGAGCACTTCAAG 1140  
Db 1081 GCCGAGGCGATGAGCCAGGCGCAACAGCGTGAACATCATGATGCAAGAGAGCACTTCAAG 1140  
QY 1141 GGGCCCCCGGCGCAACGTCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATCGCAAGAAC 1200  
Db 1141 GGGCCCCCGGCGCAACGTCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATCGCAAGAAC 1200  
QY 1201 TGCCGCGCGCCCCCGCAAGAGGGCTGCTGAAGTGGCGCAAGAGGGGCCACAGATGAAG 1260  
Db 1201 TGCCGCGCGCCCCCGCAAGAGGGCTGCTGAAGTGGCGCAAGAGGGGCCACAGATGAAG 1260  
QY 1261 GACTGCACCGAGCGCCAGGCCCACTTCTTGGGCAAGATCTGGCCAGGCCCAAGGGCGCG 1320  
Db 1261 GACTGCACCGAGCGCCAGGCCCACTTCTTGGGCAAGATCTGGCCAGGCCCAAGGGCGCG 1320

QY 1321 CCCGCAACTTCTCTGCAAGAACCGGAGCGGAGCCCGCGCCCGCCCAACCTGCCCCACCGCCCC 1380  
Db 1321 CCCGCAACTTCTCTGCAAGAACCGGAGCGGAGCCCGCGCCCGCCCAACCTGCCCCACCGCCCC 1380  
QY 1381 CCCGCGGAGAGCTTCCGCTTCCAGAGAGACCAACCCCGCGCCCGCCCAAGAGAGCCCAAGAG 1440  
Db 1381 CCCGCGGAGAGCTTCCGCTTCCAGAGAGACCAACCCCGCGCCCGCCCAAGAGAGCCCAAGAG 1440  
QY 1441 CCGGAGCCCTTACCGCGAGAGCCCTTGAACCGCCCTGCGGAGAGCTTCCGAGCGGCGCCCTG 1500  
Db 1441 CCGGAGCCCTTACCGCGAGAGCCCTTGAACCGCCCTGCGGAGAGCTTCCGAGCGGCGCCCTG 1500  
QY 1501 AGCCAGTAA 1509  
Db 1501 AGCCAGTAA 1509  
  
RESULT 2  
AX468544 1509 bp DNA linear PAT 16-JUL-2002  
LOCUS AX468544  
DEFINITION Sequence 64 from Patent WO0226209.  
ACCESSION AX468544  
VERSION AX468544.1 GI:21901374  
KEYWORDS  
SOURCE Human immunodeficiency virus 1 (HIV-1)  
ORGANISM Human immunodeficiency virus 1  
VIRUSES; Retroid viruses; Retroviridae; Lentivirus; Primate  
Lentivirus group.  
  
REFERENCE 1  
AUTHORS O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singh,M., Uimer,J. and Dubensky,T.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 64 04-APR-2002;  
CHIRON CORPORATION (US)  
FEATURES  
location/Qualifiers  
source 1..1509  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"  
  
ORIGIN  
Query Match 99.9%; Score 1507.4; DB 6; Length 1509;  
Best Local Similarity 99.9%; Pred. No. 4.7e-150;  
Matches 1508; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 ATGGGCGCGCGCGCCAGCATCTCGCGCGGAGAGAGCTGGAACAAGTGGGAGAAAGATCCGC 60  
Db 1 ATGGGCGCGCGCGCCAGCATCTCGCGCGGAGAGAGCTGGAACAAGTGGGAGAAAGATCCGC 60  
QY 61 CTGCGCGCGCGCGGCAAGAGCACTACTAGCTGAAGACCTGGTGTGGGCCAGCGCGAG 120  
Db 61 CTGCGCGCGCGCGGCAAGAGCACTACTAGCTGAAGACCTGGTGTGGGCCAGCGCGAG 120  
QY 121 CTGAGGGCTTCCGCTGAACCCCGGCTGCTGAGACCGCGGAGGGCTGCAAGCAGATC 180  
Db 121 CTGAGGGCTTCCGCTGAACCCCGGCTGCTGAGACCGCGGAGGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCACTGACGCCCGCTGACAGCCGCAACCGAGAGAGTGGCGAGCCTGTACAAC 240  
Db 181 ATGAAGCACTGACGCCCGCTGACAGCCGCAACCGAGAGAGTGGCGAGCCTGTACAAC 240  
QY 241 ACCGTGGCCACCCTGTACTGCGTGACAGCCGGCATGAGGTCCGCGACACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCCTGTACTGCGTGACAGCCGGCATGAGGTCCGCGACACCAAGAGGCC 300  
QY 301 CTGGAACAAGATCGAGGAGAGCAGAAACAAGTCCAGCAGAGAACCAGCAGGCCCAAGAG 360  
Db 301 CTGGAACAAGATCGAGGAGAGCAGAAACAAGTCCAGCAGAGAACCAGCAGGCCCAAGAG 360  
QY 361 GCCGAGCGCAAGGTGAGCCAGAACTACCTCATCGTGCAAGAACTTGACAGGGCCAGATG 420  
Db 361 GCCGAGCGCAAGGTGAGCCAGAACTACCTCATCGTGCAAGAACTTGACAGGGCCAGATG 420

```
QY 421 CACCAAGCCATCAGCCCCCGCACCCCTGAAAGCCTGGGTGAAGGTGATCGAGAGAAGGCC 480
Db 421 CACCAAGCCATCAGCCCCCGCACCCCTGAAAGCCTGGGTGAAGGTGATCGAGAGAAGGCC 480
QY 481 TTCAAGCCCCGAGGTGATCCCATGTTCAACGCCCCTGAGGAGGCGGCCACCCCGCAGGAC 540
Db 481 TTCAAGCCCCGAGGTGATCCCATGTTCAACGCCCCTGAGGAGGCGGCCACCCCGCAGGAC 540
QY 541 CTGAACAAGATGTTGAACAACCGTGGCGGCCACCAAGCGCCCATGAGATGCTGAAGGAC 600
Db 541 CTGAACAAGATGTTGAACAACCGTGGCGGCCACCAAGCGCCCATGAGATGCTGAAGGAC 600
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGGACCGCCTGCAACCCCGTGACGCGCGCGCGCTG 660
Db 601 ACCATCAACGAGAGGCGCGCGAGTGGGACCGCCTGCAACCCCGTGACGCGCGCGCGCTG 660
QY 661 GCGCGCGCGCGAGATGCGCGGACCGCGCGCGCGAGCATCGCGCGCGCGCGCGCGCGCGCTG 720
Db 661 GCGCGCGCGCGAGATGCGCGGACCGCGCGCGCGAGCATCGCGCGCGCGCGCGCGCGCGCTG 720
QY 721 CAGAGCAGATCGCTGGATGACCAAGCAACCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCTG 780
Db 721 CAGAGCAGATCGCTGGATGACCAAGCAACCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCTG 780
QY 781 CGGTGATCATCTCGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTTG 840
Db 781 CGGTGATCATCTCGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAGCATCTTG 840
QY 841 GACATCCGCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 900
Db 841 GACATCCGCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 900
QY 901 CTGCGCGCGCGAGCAGCGCCACCAAGACGTGAAGACTGATGACCGAGACCCCTGCTGCTG 960
Db 901 CTGCGCGCGCGAGCAGCGCCACCAAGACGTGAAGACTGATGACCGAGACCCCTGCTGCTG 960
QY 961 CAGAACGCCAACCCCGCACTGCAAGACATCTGCGCGCTCTCGCGCGCGCGCGCGCGCGCTG 1020
Db 961 CAGAACGCCAACCCCGCACTGCAAGACATCTGCGCGCTCTCGCGCGCGCGCGCGCGCGCTG 1020
QY 1021 GAGGAGATGATGACCGCGCTGCGAGGCGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1080
Db 1021 GAGGAGATGATGACCGCGCTGCGAGGCGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1080
QY 1081 GCGGAGGCGATGAGCGCGCAAGCGCTGAACATCATGATGCAAGAGCAACTTCAAG 1140
Db 1081 GCGGAGGCGATGAGCGCGCAAGCGCTGAACATCATGATGCAAGAGCAACTTCAAG 1140
QY 1141 GCGCGCGCGCGCAAGCGTCAAGTCTTCACTGCGCGCAAGAGGCGCAATCGCGCAAGAAC 1200
Db 1141 GCGCGCGCGCGCAAGCGTCAAGTCTTCACTGCGCGCAAGAGGCGCAATCGCGCAAGAAC 1200
QY 1201 TGGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGCAAGGAGGCGCACCAAGTGAAG 1260
Db 1201 TGGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGCAAGGAGGCGCACCAAGTGAAG 1260
QY 1261 GACTGCAACGAGCGCGCAAGCGCAACTTCTGCGCAAGATCTGCGCGCAAGGCGCGC 1320
Db 1261 GACTGCAACGAGCGCGCAAGCGCAACTTCTGCGCAAGATCTGCGCGCAAGGCGCGC 1320
QY 1321 CCGGCAACTTCTGCAAGAACCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1380
Db 1321 CCGGCAACTTCTGCAAGAACCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1380
QY 1381 CCGGCGAGAGCTTCGCTTGAAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1440
Db 1381 CCGGCGAGAGCTTCGCTTGAAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1440
QY 1441 CCGGAGCCCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1500
Db 1441 CCGGAGCCCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1500
```

```
QY 1501 AGCCAGTAA 1509
Db 1501 AGCCAGTAA 1509

RESULT 3
AX468548
LOCUS
DEFINITION
Sequence 68 from Patent WO0226209.
AX468548
VERSION
AX468548.1 GI:21901378
KEYWORDS
SOURCE
Human immunodeficiency virus 1 (HIV-1)
ORGANISM
Human immunodeficiency virus 1
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
lentivirus group.

REFERENCE
1
AUTHORS
O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,
Singh,M., Uimer,J. and Dubensky,T.W.
TITLE
Microparticles for delivery of the heterologous nucleic acids
JOURNAL
Patent: WO 0226209-A 68 04-APR-2002;
CHIRON CORPORATION (US)
FEATURES
location/Qualifiers
source
1.1509
/organism="Human immunodeficiency virus 1"
/mol_type="unassigned DNA"
/db_xref="taxon:11676"

ORIGIN

Query Match 99.7%; Score 1504.2; DB 6; Length 1509;
Best Local Similarity 99.2%; Pred.No.1e-149;
Matches 1497; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 601 ACCATCAAGAGAGGCGCCGAGTGGGACCGCCTGCAACCCCGTGCAGGCGGCGCCGCTG 660
Db 601 ACCATCAAGAGAGGCGCCGAGTGGGACCGCCTGCAACCCCGTGCAGGCGGCGCCGCTG 660
QY 661 GCGCGCGCCAGATGCGCGGACCGCCGCGGCAAGCAGATCGCGCGGCAACGACCCCTG 720
Db 661 GCGCGCGCCAGATGCGCGGACCGCCGCGGCAAGCAGATCGCGCGGCAACGACCCCTG 720
QY 721 CAGAGCAGATGCGCTGGATGACCAAGACCCCGCGTGCAGCGCGGCAATCTACAG 780
Db 721 CAGAGCAGATGCGCTGGATGACCAAGACCCCGCGTGCAGCGCGGCAATCTACAG 780
QY 781 CGGTGATCATCTGGGCTTGAACAAGATCGTGGATGTACAGCCCGGTGAGCATCTG 840
Db 781 CGGTGATCATCTGGGCTTGAACAAGATCGTGGATGTACAGCCCGGTGAGCATCTG 840
QY 841 GACATCGCCAGAGGCGCCCAAGAGCGCTTCCGCGACTACGTGAGCGCTTCTTCAAGAC 900
Db 841 GACATCGCCAGAGGCGCCCAAGAGCGCTTCCGCGACTACGTGAGCGCTTCTTCAAGAC 900
QY 901 CTGCGCGCCGAGCAGGCGCACCAAGACGTGAAGAACTGGATGACCGAGACCTGTGCTG 960
Db 901 CTGCGCGCCGAGCAGGCGCACCAAGACGTGAAGAACTGGATGACCGAGACCTGTGCTG 960
QY 961 CAGAACGCCAACCCCGACTGCAAGACATCTGCGCGCTCTCGGCGCGCGCGGCGCACCTG 1020
Db 961 CAGAACGCCAACCCCGACTGCAAGACATCTGCGCGCTCTCGGCGCGCGCGGCGCACCTG 1020
QY 1021 GAGAGATGATGACCGCGCTGCAAGGCGTGGGCGCGCGCGCGCGCACAGGCGCGCTGCTG 1080
Db 1021 GAGAGATGATGACCGCGCTGCAAGGCGTGGGCGCGCGCGCGCGCACAGGCGCGCTGCTG 1080
QY 1081 GCGGAGGCGATGAGCGGCGCAACAGCGTGAACATCATGATGACAGAGAGCAACTTCAAG 1140
Db 1081 GCGGAGGCGATGAGCGGCGCAACAGCGTGAACATCATGATGACAGAGAGCAACTTCAAG 1140
QY 1141 GCGCGCGCGCGCAAGCTCAAGTGTCTTCAACTGCGGCAAGAGGCGCACATCGCAAGAAC 1200
Db 1141 GCGCGCGCGCGCAAGCTCAAGTGTCTTCAACTGCGGCAAGAGGCGCACATCGCAAGAAC 1200
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAGGAGGCGCACAGATGAAG 1260
Db 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAGGAGGCGCACAGATGAAG 1260
QY 1261 GACTGCAACGAGCGCGCAAGCTTCTGGGCAAGATCTGCGCCAGCCACAGGCGCGC 1320
Db 1261 GACTGCAACGAGCGCGCAAGCTTCTGGGCAAGATCTGCGCCAGCCACAGGCGCGC 1320
QY 1321 CCGGCAACTTCTGCAAGACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 1321 CCGGCAACTTCTGCAAGACCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
QY 1381 CCGCGAGAGCTTCCGCTTGAAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
Db 1381 CCGCGAGAGCTTCCGCTTGAAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
QY 1441 CCGGAGCGCTTACCGCGAGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
Db 1441 CCGGAGCGCTTACCGCGAGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
QY 1501 AGCCAGTAA 1509
Db 1501 AGCCAGTAA 1509
```

```
RESULT 4
AX455905 1509 bp DNA linear PAT 06-JUL-2002
LOCUS AX455905
DEFINITION Sequence 21 from Patent WO0204493.
ACCESSION AX455905
VERSION AX455905.1 GI:21714897
KEYWORDS
SOURCE synthetic construct
```

```
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS zur Megeide,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL polypeptides and uses thereof
Patent: WO 0204493-A 21 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source 1.1509
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="synthetic Gag coding sequence of HIV strain
AF110967"
ORIGIN
Query Match 98.7%; Score 1489.8; DB 6; Length 1509;
Best Local Similarity 99.2%; Pred.No.3.4e-148;
Matches 1497; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 CTGAGGCGCTTCCGCTTGAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 121 CTGAGGCGCTTCCGCTTGAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 ATGAAGCAGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 ATGAAGCAGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 ACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 241 ACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 301 CTGGAACAAGATCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
Db 301 CTGGAACAAGATCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
QY 361 GCGGACGCGCAAGGTGAGCGCGCAACTACCCCATCTGTGCAAGAACTTGAGGGCCAGATG 420
Db 361 GCGGACGCGCAAGGTGAGCGCGCAACTACCCCATCTGTGCAAGAACTTGAGGGCCAGATG 420
QY 421 CACCAAGGCCATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 421 CACCAAGGCCATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 TTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 481 TTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 541 CTGAACAAGATGTTGAACAACCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 541 CTGAACAAGATGTTGAACAACCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
QY 601 ACCATCAAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 601 ACCATCAAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 661 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Db 661 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 721 CAGAGCAGATCGCGCTGATGACCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db 721 CAGAGCAGATCGCGCTGATGACCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
```

QY 781 CGGTGATCATCTCTGGGCTTGAACAAGATCGTCCGATGTACAGCCCCGTGAGCATCCTG 840  
Db 781 CGCTGGATCATCTCTGGGCTTGAACAAGATCGTCCGATGTACAGCCCCGTGAGCATCCTG 840  
QY 841 GACATCCGCCAGAGGCCCCCAAGAGGCCCTTCCGCGACTACGTGAACCGCTTCTTCAAGACC 900  
Db 841 GACATCCGCCAGAGGCCCCCAAGAGGCCCTTCCGCGACTACGTGAACCGCTTCTTCAAGACC 900  
QY 901 CTGCGCGCCGAGCAGAGGCCACCAGAGCGTGAAGAACTGGAATGACCCGAGACCCTGCTGTG 960  
Db 901 CTGCGCGCCGAGCAGAGGCCACCAGAGCGTGAAGAACTGGAATGACCCGAGACCCTGCTGTG 960  
QY 961 CAGAAAGCCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCGCGCCAGCCTG 1020  
Db 961 CAGAAAGCCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCGCGCCAGCCTG 1020  
QY 1021 GAGAGATGATGACCGCTGCGCAGAGGCGTGGGCGGCCCGGCCACAAGGCCCGCTGTG 1080  
Db 1021 GAGAGATGATGACCGCTGCGCAGAGGCGTGGGCGGCCCGGCCACAAGGCCCGCTGTG 1080  
QY 1081 GCCGAGCGGATGAGCCAGGCCCAACAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140  
Db 1081 GCCGAGCGGATGAGCCAGGCCCAACAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140  
QY 1141 GGCCCCCGGCGCAACGTCAAGTGTCTTCACTGCGGCAAGAGGGCCCATCGCCAAAGAAC 1200  
Db 1141 GGCCCCCGGCGCAACGTCAAGTGTCTTCACTGCGGCAAGAGGGCCCATCGCCAAAGAAC 1200  
QY 1201 TGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGAGGCCCAAGATGAAG 1260  
Db 1201 TGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGAGGCCCAAGATGAAG 1260  
QY 1261 GACTGCAACGAGCGCCAGGCCCAACTTCTGGGCAAGATCTGGCCCAAGAGGCCCGC 1320  
Db 1261 GACTGCAACGAGCGCCAGGCCCAACTTCTGGGCAAGATCTGGCCCAAGAGGCCCGC 1320  
QY 1321 CCGGCAACTTCTCTGAGAAACGCGAGGAGCCCCCGCCCAACCTGCCCCACCGCCCC 1380  
Db 1321 CCGGCAACTTCTCTGAGAAACGCGAGGAGCCCCCGCCCAACCTGCCCCACCGCCCC 1380  
QY 1381 CCGCGGAGAGCTTCGCTTTCGAGAGAACACCCCGCCCAAGCAGAGCCCAAGAAC 1440  
Db 1381 CCGCGGAGAGCTTCGCTTTCGAGAGAACACCCCGCCCAAGCAGAGCCCAAGAAC 1440  
QY 1441 CGGAGCCCTACCGGAGGCCCTGACCGCCCTGCGAGCCTGTTGCGAGCGGCCCTG 1500  
Db 1441 CGGAGCCCTACCGGAGGCCCTGACCGCCCTGCGAGCCTGTTGCGAGCGGCCCTG 1500  
QY 1501 AGCCAGTAA 1509  
Db 1501 AGCCAGTAA 1509

RESULT 5  
AX455935  
LOCUS AX455935 1494 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 51 from Patent WO0204493.  
ACCESSION AX455935  
VERSION AX455935.1 GI:21714919  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1 other sequences; artificial sequences.  
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,B.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 51 17-JAN-2002;  
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
FEATURES  
source location/Qualifiers  
1..1494  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 86.0%; Score 1297.6; DB 6; Length 1494;  
Best Local Similarity 92.7%; Pred. No. 5.9e-128;  
Matches 1402; Conservative 0; Mismatches 89; Indels 21; Gaps 3;  
/db\_xref="taxon:32630"  
/note="HIV Type C Gag optimized"

QY 1 ATGGGCGCCCGCGCGCAGCATCTTGCGCGCGCGGAGAACTGGAACAAGTGGGAGAAATCCGC 60  
Db 1 ATGGGCGCCCGCGCGCAGCATCTTGAGCGCGCGGCAAGCTGGAACAAGTGGGAGCGCATCCG 60  
QY 61 CTGCGCGCCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCGAGCCCGGAG 120  
Db 61 CTGCGCGCCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCGAGCCCGGAG 120  
QY 121 CTGAGGGGCTTGGCTTGAACCCCGGCTGTGGAAGACCGCCGAGGGCTGCAAGCAGATC 180  
Db 121 CTGAGGGGCTTGGCTTGAACCCCGGCTGTGGAAGACCGAGGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGAGCCCGCCCTGCAAGACCGGCAACGAGAGCTGCGCAGCTGTACAAC 240  
Db 181 ATGAAGCAGCTGAGCCCGCCCTGCAAGACCGGCAACGAGAGCTGCGCAGCTGTACAAC 240  
QY 241 ACCGTGGCAACCTGTGTACTGCGTGAACCGCGGCTGAGGTCCGGACACCAAGAGGCC 300  
Db 241 ACCGTGGCAACCTGTGTACTGCGTGAACCGCGGCTGAGGTCCGGACACCAAGAGGCC 300  
QY 301 CTGGAACAAGTCAGAGAGAGCAAGAACAGTCCAGCAGAAAGCCAGAGGCCAAGAG 360  
Db 301 CTGGAACAAGTCAGAGAGAGCAAGAACAGTCCAGCAGAAAGCCAGAGGCCAAGAG 360  
QY 361 GCCGACGGCAAGGTGAAGCCAGAACTACCCATCTGCAAGACTGCAAGGGCCAGATGGTG 420  
Db 361 GCCGACGGCAAGGTGAAGCCAGAACTACCCATCTGCAAGACTGCAAGGGCCAGATGGTG 420  
QY 421 CACGAGCCATCAGCCCCCGCACCTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
Db 421 CACGAGCCATCAGCCCCCGCACCTGAACGCTGGGTGAAGGTGATCGAGAGAGGCC 480  
QY 481 TTCAAGCCCGAGGTGATCCCATGTTCACCGGCTTGAAGCGAGGGGCCACCCCGAGAG 540  
Db 481 TTCAAGCCCGAGAGATCCCATGTTCACCGGCTTGAAGCGAGGGGCCACCCCGAGAG 540  
QY 541 CTGAACAGATGTTGAACACCGTGGGCGGCACAGGCGGCCATGAGATGCTGAAGAG 600  
Db 541 CTGAACAGATGCTGAACACCGTGGGCGGCACAGGCGGCCATGAGATGCTGAAGAG 600  
QY 601 ACCATCAACGAGAGGGCGCGGAGTGGAGCCGCTGACCCCGTGAAGCGCGGCCGCTG 660  
Db 601 ACCATCAACGAGAGGGCGCGGAGTGGAGCCGCTGACCCCGTGAAGCGCGGCCGCTG 660  
QY 661 GCCCGCGCGCAGATGCGGCAACCCCGCGGCAAGCATCGCGCGGCCACCAAGCACCCCTG 720  
Db 661 GCCCGCGCGCAGATGCGGCAAGCCCGCGGCAAGCATCGCGCGGCCACCAAGCACCCCTG 720  
QY 721 CAGAGCAGATCGCTGATGACCAAGCAACCCCGCTGCCCCCTGGGGCGACATTTACAAG 780  
Db 721 CAGAGCAGATCGCTGATGACCAAGCAACCCCGCTGCCCCCTGGGGCGACATTTACAAG 780  
QY 781 CGGTGATCATCTCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATCCTG 840  
Db 781 CGGTGATCATCTCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATCCTG 840  
QY 841 GACATCCGCCAGAGGCCCCCAAGAGGCCCTTCCGCGACTACGTGAACCGCTTCTTCAAGACC 900  
Db 841 GACATCAAGCAGAGGCCCCCAAGAGGCCCTTCCGCGACTACGTGAACCGCTTCTTCAAGACC 900  
QY 901 CTGCGCGCGGAGCAGGCCCAACCAAGAGCGTGAAGAACTGATGACCGAGACCCTGCTGTG 960  
Db 901 CTGCGCGCGGAGCAGGCCCAACCAAGAGCGTGAAGAACTGATGACCGAGACCCTGCTGTG 960  
QY 961 CAGAAAGCCCAACCCCGACTGCAAGACCATCTGCGGCTCTGCGGCCCGCGCCACCCCTG 1020

Db 961 CAGAACGCCAACCCCGACTGCAAGACCACTCTGCGCGCCCTGGGGCCCGCGCCAGCCTTG 1020

QY 1021 GAGAGATGATGACCGCTTCCAGGGCGCTGGGGCGCCCGCCAGAGGCCCGCTGCTG 1080

Db 1021 GAGAGATGATGACCGCTTCCAGGGCGCTGGGGCGCCCGCCAGAGGCCCGCTGCTG 1080

QY 1081 GCCGAGCGCATGAGCCAGGCCAACAGCGCTGAAATCATGATGCAAGAGCAACTTCAAG 1140

Db 1081 GCCGAGCGCATGAGCCAGGCCAACAGCGCTGAAATCATGATGCAAGAGCAACTTCAAG 1137

QY 1141 GCGCCCGCGCGCAACGTCAAGTGTCTCACTGCGGCAAGAGGGCCACATCGCCAAAGAAC 1200

Db 1138 GCGAGCAACCGCATCATAGTGTCTCACTGCGGCAAGTGGGCCACATCGCCGCAAC 1197

QY 1201 TGCCGCGCCCGCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACCATGTAAG 1260

Db 1198 TGCCGCGCCCGCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACCATGTAAG 1257

QY 1261 GACTGCAACCGAGCGCGCAAGGCTGCTGGAAGTGGCGCAAGATCTGGCCCAAGAGGGCCG 1320

Db 1258 GACTGCAACCGAGCGCGCAAGGCTGCTGGAAGTGGCGCAAGATCTGGCCCAAGAGGGCCG 1317

QY 1321 CCGCGCACTTCTGCAAGACCGAGCGAGC--CCGCGCGCCCGCCACCGTGCCCAACCGCC 1377

Db 1318 CCGCGCACTTCTGCAAGACCGCGAGCGCCCGCCCGCCCGCGCGAGCCCGCC 1377

QY 1378 CCGCGCGCGAGAGCTCCGCTTGAAGAGACCAACCGCCCGCCCGCCAGAGAGCCCAAG 1437

Db 1378 CCGCGCGCGAGAGCTCCGCTTGAAGAGACCAACCGCCCGCCCGCCAGAGAGAGAG 1437

QY 1438 GACCGGAGCGCTACCGCGAGCGCCCTGACCGCCCTGCGCAAGCTGTTCCGAGCGCGCC 1497

Db 1438 -----CGCGAGCGCCCTGACCAAGCTGTAAGAGCTGTTCCGAGCGAGCCCGCC 1482

QY 1498 CTGAGCCAGTAA 1509

Db 1483 CTGAGCCAGTAA 1494

RESULT 6  
AX455983  
LOCUS AX455983 1491 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 99 from Patent WO0204493.  
ACCESSION AX455983  
VERSION AX455983.1 GI:21714967  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
ORIGIN

1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
Polynucleotides encoding antigenic hiv type c polypeptides,  
Patent: WO 0204493-A 99 17-JAN-2002;  
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
location/Qualifiers  
1. 1491  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Gag\_TV2\_C\_ZAopt"

Query Match 85.8%; Score 1294.8; DB 6; Length 1491;  
Best Local Similarity 92.7%; Pred. No. 1.2e-127;  
Matches 1401; Conservative 0; Mismatches 87; Indels 24; Gaps 3;

QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGCGAGAGCTGCAAGTGGAGAAATCCGC 60  
Db 1 ATGGGCGCGCGCGCGCATCTGCGCGCGCGAGAGCTGCAAGTGGAGAAATCCGC 60  
QY 61 CTGCGCGCGCGCGCGCAAGAGCACTAATGCTGAAGCACTGTGTGGGCCAGCGCGAG 120

Db 61 CTGCGCGCGCGCGCGCAAGCACTAATGCTGAAGCACTGTGTGGGCCAGCGCGAG 120

QY 121 CTGGAAGGCTTCCGCTTGAACCCCGCGCTGCTGGAAGACCGCGGAGGGCTGCAAGCATC 180

Db 121 CTGGAAGGCTTCCGCTTGAACCCCGCGCTGCTGGAAGACCGCGGAGGGCTGCAAGCATC 180

QY 181 ATGAAGCAGCTGAGCGCCCGCTGCAAGACCGGCAACCGAGAGCTGCGAGCCTGTACAC 240

Db 181 ATGAAGCAGCTGAGCGCCCGCTGCAAGACCGGCAACCGAGAGATCCGAGCCTGTACAC 240

QY 241 ACCGTGGCCACCTGTACTGCGGTGCAAGCGCGCATCGAGGTCCGCGACACCAAGAGGCC 300

Db 241 ACCGTGGCCACCTGTACTGCGGTGCAAGCGCGCATCGAGGTCCGCGACACCAAGAGGCC 300

QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCCAAGCAAGAACCCAGCGGCCAAGAG 360

Db 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCCAAGCAAGAACCCAGCGGCCAAGAG 360

QY 361 GCGGACGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAAGACCTGCAAGGGCCAGATG 420

Db 361 GCGGACGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAAGACCTGCAAGGGCCAGATG 420

QY 421 CACCAAGGCGATCAGCCCGCGCACCTGAAAGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480

Db 421 CACCAAGGCGATCAGCCCGCGCACCTGAAAGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480

QY 481 TTCAAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCGAGAG 540

Db 481 TTCAAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCGAGAG 540

QY 541 CTGAACAAGATGTTGAACAACCGTGGCGGCGCAAGCGCGCATGCAAGTCTGAAGAGAC 600

Db 541 CTGAACAAGATGTTGAACAACCGTGGCGGCGCAAGCGCGCATGCAAGTCTGAAGAGAC 600

QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGACCCCGCTGACGCGCGCGCCCGT 660

Db 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGACCCCGCTGACGCGCGCGCCCGT 660

QY 661 GCGCGCGCGCGAGATGCGCGACCCCGCGCGAGCGACATCGCCGCGCCACCAAGCAACCT 720

Db 661 GCGCGCGCGCGAGATGCGCGACCCCGCGCGAGCGACATCGCCGCGCCACCAAGCAACCT 720

QY 721 CAGGAGCAAGTCCGCTGATGACCAAGCAACCCCGCTGCGCGCGCGCATCTACAAG 780

Db 721 CAGGAGCAAGTCCGCTGATGACCAAGCAACCCCGCTGCGCGCGCGCATCTACAAG 780

QY 781 CGGTGATCATCTTGGGCTGAAACAAGATGCTGGGATGTAACGCCCGTGAACATCTCTG 840

Db 781 CGGTGATCATCTTGGGCTGAAACAAGATGCTGGGATGTAACGCCCGTGAACATCTCTG 840

QY 841 GACATCCGCGAGGGCGCCCAAGAGCGCTTCCGCGAGCTGGAACCGCTTCTCAAGACC 900

Db 841 GACATCCGCGAGGGCGCCCAAGAGCGCTTCCGCGAGCTGGAACCGCTTCTCAAGACC 900

QY 901 CTGCGCGCGCGAGAGCGCCCAAGAGCGTGAAGAACTGGAATGACCGAGACCTGTGCTG 960

Db 901 CTGCGCGCGCGAGAGAGCGCCCAAGAGCGTGAAGAACTGGAATGACCGAGACCTGTGCTG 960

QY 961 CAGAACCGCAACCCCGCATGCAAGACCATCTGCGCGCTCTCGGCGCGCGCGCCACCTG 1020

Db 961 CAGAACCGCAACCCCGCATGCAAGACCATCTGCGCGCTCTCGGCGCGCGCGCCACCTG 1020

QY 1021 GAGGAGATGATGACCGCTTCCAGAGGGCGTGGGGCGCCCGCCAGAGCGCGCTGCTG 1080

Db 1021 GAGGAGATGATGACCGCTTCCAGAGGGCGTGGGGCGCCCGCCAGAGCGCGCTGCTG 1080

QY 1081 GCCGAGGCGATGAGCGCGCAAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140

Db 1081 GCCGAGGCGATGAGCGCGCAAGCGTGAACATCATGATGCAAGAGCAACTTCAAG 1140

QY 1141 GCGCGCGCGCGCAAGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCAAAGAAC 1200

Db 1141 GGGCCCCCGCGCGCGTGAAGTGCTTCAACTGCGGCGCGGAGGGCCACATCGCCCCGCAAC 1200

Qy 1201 TGCCGCGCGCCCCCGGCAAGAGGCTGTGGAAGTGCAGCAAGAGGGCCACCATGATGAAG 1260

Db 1201 TGCCGCGCGCCCCCGGCAAGCGCGCTGTGGAAGTGCAGCAAGAGGGCCACCATGATGAAG 1260

Qy 1261 GACTGCACCGAGCGCGGCAAGGCTTCTGGGCAAGATCTGGCCACGACCAAGAGGGCGCC 1320

Db 1261 GACTGCACCGAGCGCGGCAAGGCTTCTGGGCAAGATCTGGCCACGACCAAGAGGGCGCC 1320

Qy 1321 CCGGCAACTTCTGTGAGAACCGCAGCGAGCCCGCGCGCCCCCACC---GTGCCCCACCGCC 1377

Db 1321 CCGGCAACTTCTGTGAGAACCGCGCCGAGCCCAACCGCCCCCTGTGAGCCACCGCC 1380

Qy 1378 CCGGCGCGCGGAGAGCTTCCGCTTCAAGAGACCAACCCCGCCCCCAAGAGAGCCCAAG 1437

Db 1381 CCGGCGCGCGGAGAGCTTCAAGTTCAGAGAGA-----CCCCCAAGAGAGAGCCCAAG 1431

Qy 1438 GACCGCGAGCCCTAACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTCGGCAGCGCGCC 1497

Db 1432 G-----ACCGCGAGCCCTGACCAAGCCTGTTCGGCAGCGACCCC 1479

Qy 1498 CTGAGCCAGTAA 1509

Db 1480 CTGAGCCAGTAA 1491

RESULT 7

AY181195 1485 bp DNA linear SYN 02-OCT-2003

LOCUS AY181195

DEFINITION Synthetic construct HIV-1-derived gag protein (gag) gene, complete cds.

ACCESSION AY181195

VERSION AY181195.1 GI:37413992

KEYWORDS

SOURCE synthetic construct

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 1485)

AUTHORS Gao,F., Li,Y., Decker,J.M., Peyerl,F.W., Bibollet-Ruche,F., Rodenburg,C.M., Chen,Y., Shaw,D.R., Allen,S., Musonda,R., Shaw,G.M., Zaijac,A.J., Letvin,N. and Hahn,B.H.

TITLE Codon usage optimization of HIV type 1 subtype C gag, pol, env, and nef genes: in vitro expression and immune responses in DNA-vaccinated mice

JOURNAL AIDS Res. Hum. Retroviruses 19 (9), 817-823 (2003)

REFERENCE 2 (bases 1 to 1485)

AUTHORS Gao,F., Li,Y., Decker,J.M., Bibollet-Ruche,F., Rodenburg,C.M., Chen,Y., Shaw,D.R., Allen,S., Musonda,R., Shaw,G.M., Zaijac,A.J. and Hahn,B.H.

TITLE Direct Submission

JOURNAL Submitted (16-NOV-2002) Medicine, Duke University Medical Center, 112 RPIII, Lasalle Street, Durham, NC 27710, USA

FEATURES

source

1. .1485

/organism="synthetic construct"

/mol\_type="genomic DNA"

/db\_xref="taxon:32630"

/note="codon usage optimization of HIV-1 96ZM651"

1. .1485

/gene="gag"

1. .1485

/gene="gag"

/codon\_start=1

/transl\_table=1

/product="gag protein"

/protein\_id="AA065393.1"

/db\_xref="GI:37413993"

/translation="MGARASILRGKLDKWEKIRLRPGSKRYMIKHLVWASRELERF

ALNPGLETSEGCKOIMKQIPALQGTBELRSLYNTVAITLYCVHEGVVDRDTKEALD

RIEBSQNKIQOKIQKQTQAADGKVSQNPPIVONLOGVNHOKLSPTLINAMVKVIEE

KAFSPFVIMFTALSEGATPDILNTMLNTVGHQAAMQMLKDTINEBAEWDRLHPVH

AGPIAPGQMBEPRGSDIAGTSTLQEQIAMWTSNPPIPVGDIYKRWIILGLNKIVRMV

SPVSLIDIKQPKPEPRDYVDRFFFKTLRAEQATQEVKNMTDTLLVQANPDCKTILK

ALPGATLEMBMTACOGVGNVSHKARVLAEMSQTSNVILIMOKSNFKRMVKCFN

CGKEGHIAARNCAPRKRCWKCGKEGHQMKDCTERQANFLGIWPSHKRPGNFLONR

PEPTAPPAESFRFEETTPAPKQESKDREALTSLKSLFSGDPLSQ"

ORIGIN

Query Match 85.6%; Score 1291.4; DB 12; Length 1485;

Best Local Similarity 92.6%; Pred. No. 2.7e-127;

Matches 1403; Conservative 0; Mismatches 76; Indels 36; Gaps 3;

Qy 1 ATGGGCGCGCGCGCGAGCATCTCTGGCGGGCGAGAACTGGAACAAGTGGGAGAAGATCCGC 60

Db 1 ATGGGCGCGCGCGCGAGCATCTCTGGCGGGCGAGAACTGGAACAAGTGGGAGAAGATCCGC 60

Qy 61 CTGGCGCGCGCGCGCAAGACATCATGTGAAGCACTGTGTGGCCAGCCGCGAG 120

Db 61 CTGGCGCGCGCGCGCAAGACGCTTACATGATCAAGCACTGTGTGGCCAGCCGCGAG 120

Qy 121 CTGGAGGGCTTCGCTTGAACCCCGGCTGTGAGAACCGCGAGGGCTGCAAGCAGATC 180

Db 121 CTGGAGCGCTTCGCTTGAACCCCGGCTGTGAGAACCGCGAGGGCTGCAAGCAGATC 180

Qy 181 ATGAAGCAGCTGCAGCGCGCGCTGAGACCGGCAACGAGAGCTGCGCAGCCTGTACAAC 240

Db 181 ATGAAGCAGCTGCAGCGCGCGCTGAGACCGGCAACGAGAGCTGCGCAGCCTGTACAAC 240

Qy 241 ACCGTGGCCACCCTGTACTGCGTGCACGCGGCAATCGAGGTCGCGGACCAAGAGGCC 300

Db 241 ACCGTGGCCACCCTGTACTGCGTGCACGAGGCGTGAAGGTGCGGACCAAGAGAGGCC 300

Qy 301 CTGGACAAGATCGAGAGAGAGAGACAAGTCCAGCAGAAGACCAGAGGCCA----- 355

Db 301 CTGGACCGCATCGAGAGAGAGAGACAAGATCCAGCAGAAGATCCAGAAAACCCAG 360

Qy 356 -AGGAGCGCGAGCGCAAGGTGAGCGCAGAACTACCCCATCGTGACAGAGGCCAG 414

Db 361 CAGGCGCGCGAGCGCAAGGTGAGCGCAGAACTACCCCATCGTGACAGAGGCCAG 420

Qy 415 ATGTGTACACGAGGCTACGCCCCCGCACCTGAAAGCCTGGGTGAAGGTGATCGAGAG 474

Db 421 ATGTGTACACGAGAGCTGAGCCCGCGCACCTGAAAGCCTGGGTGAAGGTGATCGAGAG 480

Qy 475 AAGGCTTACGCGCGGAGGTATCCCATGTTCAACCGCCCTGAGCGAGGGCGCACCCCC 534

Db 481 AAGGCTTACGCGCGGAGGTATCCCATGTTCAACCGCCCTGAGGAGGGCGCACCCCC 540

Qy 535 CAGGACCTGAACAGATGTTGAACACCGTGGCGGCCACGAGCGCCATGACATGCTG 594

Db 541 CAGGACCTGAACACATGCTGAACACCGTGGCGGCCACGAGCGCCATGACATGCTG 600

Qy 595 AAGGACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGACCCCGTGAGGCGGCG 654

Db 601 AAGGACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGACCCCGTGACGCGGCG 660

Qy 655 CCGGTGGCGCGCGCGAGATGCGCGACCCCGCGGAGCGACATGCGCGGCGCACAGC 714

Db 661 CCGATCGCGCGCGCGAGATGCGCGAGCCCGCGGAGCGACATGCGCGGCGCACAGC 720

Qy 715 ACCCTGCAGAGAGATCGCCTGATGACCAAGCAACCCCGTGCCTGGCGGACATC 774

Db 721 ACCCTGCAGAGAGATCGCCTGATGACCAAGCAACCCCGTGCCTGGCGGACATC 780

Qy 775 TACAAGCGGTGATCATCTGGGCTGAACAAGATCGTGCAGATGTACAGCCCGTGAGC 834

Db 781 TACAAGCGGTGATCATCTGGGCTGAACAAGATCGTGCAGATGTACAGCCCGTGAGC 840

Qy 835 ATCTTGACATCCGCCAGAGGCGCCCAAGAGCCCTTCCGCGACTACGTGAGACCGCTTCTC 894

Db 841 ATCTTGACATCAAGAGGCGCCCAAGAGCCCTTCCGCGACTACGTGAGACCGCTTCTC 900

Qy 895 AAGACCTGCGCGCGCGAGAGGCCACCAAGAGCTGAAGACTGACCGAGACCCCTG 954

Db 901 AAGACCTGCGCGCGCGAGAGGCCACCAAGAGCTGAAGACTGACCGAGACCCCTG 960

QY 955 CTGTGAGAAAGCCAAACCCGAGTGAAGACCATCTGCGGCTCTGCGCCCCCGGCGC 1014  
Db 961 CTGTGAGAAAGCCAAACCCGAGTGAAGACCATCTGGAAGCCCTGGGCCCCGCGCGC 1020  
QY 1015 ACCCTGAGAGATGATGACCGGCTGCGAGGGCGTGGGCGGCCCCGCGCAAGGCCCGC 1074  
Db 1021 ACCCTGAGAGATGATGACCGGCTGCGAGGGCGTGGGCGGCCCCGCGCAAGGCCCGC 1080  
QY 1075 GTGCTGCGCGAGCGATGAGCCAGCCAAAGCGTGAACATCATGATGAGAAAGCAAC 1134  
Db 1081 GTGCTGCGCGAGCGATGAGCCAGCCAAAGCGTGAACATCATGATGAGAAAGCAAC 1140  
QY 1135 TTCAAGGCGCCCGCGCGCAAGTGAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCC 1194  
Db 1141 TTCAAGGCGCAAGCGCATGTGAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCC 1200  
QY 1195 AAGAACTGCGCGCGCCCGCGCAAGAGGGCTGTGAAGTGCAGAGAGGGCCACAG 1254  
Db 1201 CGCAACTGCGCGCGCCCGCGCAAGAGGGCTGTGAAGTGCAGAGAGGGCCACAG 1260  
QY 1255 ATGAAGACTGCACCGAGCGCCAGGCCAATTCTGGGCAAGATCTGGCCAGCAAG 1314  
Db 1261 ATGAAGACTGCACCGAGCGCCAGGCCAATTCTGGGCAAGATCTGGCCAGCAAG 1320  
QY 1315 GGCGCGCGCGCAACTTCTGAGAAAGCGCAGCGCGCGCGCCCGCCACCGTGCCAGC 1374  
Db 1321 GGCGCGCGCGCAACTTCTGAGAAAGCGC-----CGCCGAGCCAGC 1362  
QY 1375 GCGCGCGCGCGCGAGAGCTTCCGCTTGAAGAGAGACACCGCGCGCCCAAGCAGAGCC 1434  
Db 1363 GCGCGCGCGCGCGAGAGCTTCCGCTTGAAGAGAGACACCGCGCGCCCAAGCAGAGAGC 1422  
QY 1435 AAGAACCGCGAGCCCTAACCGGAGCCCTGACCGCGCTGCGAGCGCTTGGCAGCGGC 1494  
Db 1423 AAGG-----ACCGGAGCGCCTGACCAAGCCTGAAGAGCGCTTGGCAGCGAGC 1470  
QY 1495 CCCCTGAGCCAGTAA 1509  
Db 1471 CCCCTGAGCCAGTAA 1485

RESULT 8  
AX455887 1479 bp DNA linear PAT 06-JUL-2002  
LOCUS AX455887  
DEFINITION Sequence 3 from Patent WO0204493.  
ACCESSION AX455887  
VERSION AX455887.1 GI:21714880  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1 other sequences; artificial sequences.  
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
JOURNAL polypeptides and uses thereof  
Patent: WO 0204493-A 3 17-JAN-2002;  
FEATURES CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
source location/Qualifiers  
1. 1479  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="synthetic gag of HIV strain AF110965"

ORIGIN  
Query Match 84.6%; Score 1276.8; DB 6; Length 1479;  
Best Local Similarity 92.5%; Pred. No. 9.2e-126;  
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;

QY 1 ATGGGCGCCCGCGCCAGCATCTTGGCGGCGGCGAGAGCTGACAAAGTGGAGAAATCCGC 60  
Db 1 ATGGGCGCCCGCGCCAGCATCTTGGCGGCGGCGAGAGCTGACAAAGTGGAGCGCATCCGC 60

QY 61 CTGCGCCCCCGCGCAAGAGCACTACATGCTGAAGCAACCTGTGTGGCCAGCCGCGAG 120  
Db 61 CTGCGCCCCCGCGCAAGAGTGTCTACATGATGAAGCACTGTGTGTGGCCAGCCGCGAG 120  
QY 121 CTGGAAGGGCTTCGCTCTGAACCCCGGCGCTGCTGAAGACCGCCGAGGGCTGCAAGCATC 180  
Db 121 CTGGAAGAGTTCGCTCTGAACCCCGGCGCTGCTGAAGACCGCCGAGGGCTGCAAGCATC 180  
QY 181 ATGAAGCAGCTGCAGCCCGCGCTGCAAGCCGCGCACCGAGAGCTGCGCAGCTGTCAAC 240  
Db 181 ATCGCCAGCTGCAGCCCGCGCTGCAAGCCGCGCACCGAGAGCTGAAAGACCTGTCAAC 240  
QY 241 ACCGTGGCCACCTCTGTACTGCGTGCACGCGCGCATCGAGGTCCGCGACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTCTGTACTGCGTGCACGAGAGATCGAGGTCCGCGACCAAGAGGCC 300  
QY 301 CTGGAACAAGATCGAAGAGAGAGCAACAAGTCCCAAGCAAGACCCAGCGCCAGAG 360  
Db 301 CTGGAACAAGATCGAAGAGAGAGCAACAAGTCCCAAGCAAGATCCAGCAGCGCCAGAG 360  
QY 361 GCGGAC--GCGAAGGTGAGCCGAACCTACCCCATCTGTGCAAGACCTGAGGGCCAGATG 417  
Db 361 GCGGACAAAGGCGAAGGTGAGCCGAACCTACCCCATCTGTGCAAGACCTGAGGGCCAGATG 420  
QY 418 GTGCACCAAGGCCATCAGCCCCCGCACCTGTAAACGCTGGGTGAAGGTATCGAAGAGAG 477  
Db 421 GTGCACCAAGGCCATCAGCCCCCGCACCTGTAAACGCTGGGTGAAGGTATCGAAGAGAG 480  
QY 478 GCCTTCAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGCCACCCCGAG 537  
Db 481 GCCTTCAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGCCACCCCGAG 540  
QY 538 GACCTGAACAAGATGTTGAACAACCGTGCGCGCCCAAGGCGCCCATGCAATGCTGAAG 597  
Db 541 GACCTGAACAAGATGTTGAACAACCGTGCGCGCCCAAGGCGCCCATGCAATGCTGAAG 600  
QY 598 GACACCATCAACGAGAGGCGCGCGAGTGCGGACCGCGTGCACCCCGTGCAAGCGCGCC 657  
Db 601 GACACCATCAACGAGAGGCGCGCGAGTGCGGACCGCGTGCACCCCGTGCAAGCGCGCC 660  
QY 658 GTGCGCGCGCGCGAGATGCGCGACCCCGCGCGAGCGACATCGCGCGCCCAAGCAGCACC 717  
Db 661 ATCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGCGACATCGCGCGCCCAAGCAGCACC 720  
QY 718 CTGCAGAGAGATGCGCGTGAATGACCAAGCAACCCCGCGTGGCGCGCATCTTAC 777  
Db 721 CTGCAGAGAGATGCGCGTGAATGACCAAGCAACCCCGCATCTCCGTGGCGCATCTTAC 780  
QY 778 AAGCGGTGATCATCTTGGGCTGAACAAGATGCTGCGGATGTACAGCCCGGTGAGCATC 837  
Db 781 AAGCGGTGATCATCTTGGGCTGAACAAGATGCTGCGGATGTACAGCCCGGTGAGCATC 840  
QY 838 CTGACATCCGCGCAGGCGCCCAAGAGAGCGCTTCCGCGACTACGTGAGCCGCTTCTTAAG 897  
Db 841 CTGACATCAAGCAGGCGCCCAAGAGAGCGCTTCCGCGACTACGTGAGCCGCTTCTTAAG 900  
QY 898 ACCCTGCGCGCGAGCAGCGCCACCAAGAGCTGAAGACTGATGACCAAGACCTGCTG 957  
Db 901 ACCCTGCGCGCGAGCAGCGCCACCAAGAGCTGAAGACTGATGACCAAGACCTGCTG 960  
QY 958 GTGCAGAACGCCAACCCCGCACTGCAAGACCATCTTGGCGGCTCTCGGCGCGCGCCAGC 1017  
Db 961 GTGCAGAACGCCAACCCCGCACTGCAAGACCATCTTGGCGGCTCTCGGCGCGCGCCAGC 1020  
QY 1018 CTGAGAGATGATGACCGCTTCCAGAGGCGTGGCGGCGCGCCCAAGAGCGCGCGT 1077  
Db 1021 CTGAGAGATGATGACCGCTTCCAGAGGCGTGGCGGCGCGCCCAAGAGCGCGCGT 1080  
QY 1078 CTGCGCGAGCGGATGAGCGCAAGCGCAAGCGGTGAACATCATGATGCAAGAGCAACTTC 1137  
Db 1081 CTGCGCGAGCGGATGAGCGCAAGCGCAAGCGGTGAACATCATGATGCAAGAGCAACTTC 1137

QY 1138 AAGGCCCCCGGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCAG 1197  
 DB 1138 AAGGCCCCCGGCGCATCTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCGC 1197  
 QY 1198 AACTGCGCGCGCCCCCGGCAAGAGGGCTGTGGAAGTGCAGGCAAGAGGGCCACAGATG 1257  
 DB 1198 AACTGCGCGCGCCCCCGGCAAGAGGGCTGTGGAAGTGCAGGCAAGAGGGCCACAGATG 1257  
 QY 1258 AAGGACTGCACCGAGCGCCAGGCCAACTTCTGAGCAAGATCTGAGCCAGCCACAGAGGC 1317  
 DB 1258 AAGGACTGCACCGAGCGCCAGGCCAACTTCTGAGCAAGATCTGAGCCAGCCACAGAGGC 1317  
 QY 1318 CGCCCCGGCAACTTCTGTGCAGAAACCGAGCGAGCCCGCGCCCCCACCGTGCCACCGCC 1377  
 DB 1318 CGCCCCGGCAACTTCTGTGCAGAGCCG-----CGCCGAGCCACCGCC 1359  
 QY 1378 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCCCCCAAGAGAGCCCAAG 1437  
 DB 1360 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCCCCCAAGAGAGAGCAAG 1419  
 QY 1438 GACCGCGAGCCCTTACCGCGAGCCCTGTGACCGCCCTGCGCAGCCTGTTCGGCAGCGGCC 1497  
 DB 1420 G-----ACCGCGAGACCTGTGACCGCCTGTGAAGAGCCTGTTCGGCAACGACCC 1467  
 QY 1498 CTGAGCCAGTAA 1509  
 DB 1468 CTGAGCCAGTAA 1479

RESULT 9  
 AX468543 1479 bp DNA linear PAT 16-JUL-2002  
 LOCUS AX468543 Sequence 63 from Patent WO0226209.  
 DEFINITION AX468543  
 ACCESSION AX468543  
 VERSION AX468543.1 GI:21901373  
 KEYWORDS  
 SOURCE Human immunodeficiency virus 1 (HIV-1)  
 ORGANISM Human immunodeficiency virus 1  
 Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate  
 lentivirus group.  
 REFERENCE  
 1 O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
 Singh,M., Ulmer,J. and Dubensky,T.W.  
 TITLE Microparticles for delivery of the heterologous nucleic acids  
 JOURNAL Patent: WO 0226209-A 63 04-APR-2002;  
 CHIRON CORPORATION (US)  
 FEATURES  
 source 1..1479  
 /organism="Human immunodeficiency virus 1"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:11676"

ORIGIN

Query Match 84.6%; Score 1276.8; DB 6; Length 1479;  
 Best Local Similarity 92.5%; Pred. No. 9.2e-126;  
 Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;  
 QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGGAGAGCTGAGCAAGTGGAGAGATCCGC 60  
 DB 1 ATGGGCGCGCGCGCGCATCTGCGCGCGGAGAGCTGAGCAAGTGGAGAGATCCGC 60  
 QY 61 CTGCGCGCGCGCGCGCAAGAGCACTATGCTGAAGCACTGTGTGGCCAGCGCGAG 120  
 DB 61 CTGCGCGCGCGCGCGCAAGAGTGTATGATGAGCACTGTGTGGCCAGCGCGAG 120  
 QY 121 CTGGAAGGCTTCCGCTTGAACCCCGGCTGTGAGAGCCGCGAGGGCTGCAAGCATC 180  
 DB 121 CTGGAAGAGTTCCGCTTGAACCCCGGCTGTGAGAGCCGCGAGGGCTGCAAGCATC 180  
 QY 181 ATGAAGCAGCTGCAGCGCGCGCTGAGAGCCGCGCAGAGAGCTGCGCAGCTGTACAAC 240  
 DB 181 ATCCGCCAGCTGCAGCGCGCGCTGAGAGCCGCGCAGAGAGCTGAAGAGCTGTACAAC 240

QY 241 ACCGTGGCCACCCCTGTATCTGCGTGACCGCCGCGATCGAGGTCCGCGACACCAAGAGGCC 300  
 DB 241 ACCGTGGCCACCCCTGTATCTGCGTGACGAGAAAGATCGAGGTCCGCGACACCAAGAGGCC 300  
 QY 301 CTGGAACAAGATCGAGAGAGAGCAGAAACAAGTCCACAGAGAGACCGAGGCCAAGAG 360  
 DB 301 CTGGAACAAGATCGAGAGAGAGCAGAAACAAGTCCACAGAGAGATCGAGGCCAAGAGCC 360  
 QY 361 GCCGAC--GGCAAGTGAAGCCAGAACTACCCCATCTGTCAAGAACTGTCAAGGGCCAGATG 417  
 DB 361 GCCGACAAAGGGCAAGGTGAGCCAGAACTACCCCATCTGTCAAGAACTGTCAAGGGCCAGATG 420  
 QY 418 GTGACCAAGGCCATCAGCCCCCGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAA 477  
 DB 421 GTGACCAAGGCCATCAGCCCCCGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAA 480  
 QY 478 GCCTTACGCCCGGAGGTGATCCCATGTTCACCGCCTGAGCGAGGGCCACCCCCAG 537  
 DB 481 GCCTTACGCCCGGAGGTGATCCCATGTTCACCGCCTGAGCGAGGGCCACCCCCAG 540  
 QY 538 GACCTGAACAAGATGTGAACAACGTGGCGGCGCAACAGCGCCCATGCAAGTGTGAAG 597  
 DB 541 GACCTGAACAAGATGTGAACAACGTGGCGGCGCAACAGCGCCCATGCAAGTGTGAAG 600  
 QY 598 GACACCATCAACGAGAGAGCGCGCGAGTGGACCGCCTGACCCCGTGACAGCGCGGCC 657  
 DB 601 GACACCATCAACGAGAGAGCGCGCGAGTGGACCGCCTGACCCCGTGACAGCGCGGCC 660  
 QY 658 GTGCCCCCGGCGAGATGCGCGGACCCCGCGCGAGCATCGCGCGCCACAGCAACC 717  
 DB 661 ATGCCCCCGGCGAGATGCGCGGAGCCCGCGCGAGCATCGCGCGCCACAGCAACC 720  
 QY 718 CTGCAAGAGCAGATCGCTGTGATGACCAAGACCCCGCGTGCCTGGCGGCACTTAC 777  
 DB 721 CTGCAAGAGCAGATCGCTGTGATGACCAAGACCCCGCGTGCCTGGCGGCACTTAC 780  
 QY 778 AAGCGGTGATCATCTTGGGCTGAACAAGATCGTGGAGTGTACAGCCCGTGAGCATC 837  
 DB 781 AAGCGGTGATCATCTTGGGCTGAACAAGATCGTGGAGTGTACAGCCCGTGAGCATC 840  
 QY 838 CTGACATCCGCGAGGGGCCCAAGAGGCCCTTCCGCGACTACGTGAGCCGCTTCAAG 897  
 DB 841 CTGACATCAAGAGGGGCCCAAGAGGCCCTTCCGCGACTACGTGAGCCGCTTCAAG 900  
 QY 898 ACCCTGCGCGCGGAGCAGGCCAAGCAAGACGTGAAGAACTGATGACCGAGACCTGTCTG 957  
 DB 901 ACCCTGCGCGCGGAGCAGGCCAAGCAAGACGTGAAGAACTGATGACCGAGACCTGTCTG 960  
 QY 958 GTGCAAGACGCCAACCCCGCATGCAAGACCATCTGCGCGCTTCCGCGCCCGCCACC 1017  
 DB 961 GTGCAAGACGCCAACCCCGCATGCAAGACCATCTGCGCGCTTCCGCGCCCGCCACC 1020  
 QY 1018 CTGGAAGAGATGATGACCGCTGCGCAAGGGCGTGGCGGGCCCCCGGCAAGAGCCCGCTG 1077  
 DB 1021 CTGGAAGAGATGATGACCGCTGCGCAAGGGCGTGGCGGGCCCCCGGCAAGAGCCCGCTG 1080  
 QY 1078 CTGCGCGAGGCGCATGAGCCAGGCGCAACAGCTGAACATCATGATGCAAGAGCACTTC 1137  
 DB 1081 CTGCGCGAGGCGCATGAGCCAGGCGCAACAC--AGCGTATGATGCAAGAGCACTTC 1137  
 QY 1138 AAGGGCCCCCGGCGCAAGTGAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCAG 1197  
 DB 1138 AAGGGCCCCCGGCGCATCTGTAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGC 1197  
 QY 1198 AACTGCGCGCGCCCCCGCAAGAGGGCTGTGAGAGTGCAGAGAGGGCCACAGATG 1257  
 DB 1198 AACTGCGCGCGCCCCCGCAAGAGGGCTGTGAGAGTGCAGAGAGGGCCACAGATG 1257  
 QY 1258 AAGGACTGCACCGAGCGCCAGGCCAATTCTGTGGCAAGATCTGGCCCAAGAGGGC 1317  
 DB 1258 AAGGACTGCACCGAGCGCCAGGCCAATTCTGTGGCAAGATCTGGCCCAAGAGGGC 1317  
 QY 1318 CGCCCCGGCAACTTCTGTCAAGAACCGGAGCGCGCGCCCCCAACCGTGCCACCGCC 1377

Db 1318 CGCCCCGGAAGTCTCTGAGAGCCG-----CCCGAGCCCAAGCCG 1359  
QY 1378 CCCCCCGGAGAGCTTCGGCTTGAGAGAGACACCCCGCCCCCAAGAGAGAGCCCAAG 1437  
Db 1360 CCCCCCGGAGAGCTTCGGCTTGAGAGAGACACCCCGCCCCAAGAGAGAGAGCAAG 1419  
QY 1438 GACCGGAGAGCCCTACCGGAGAGCCCTGAACCGCCCTGCGAGAGCTGTTCCGAGAGCCG 1497  
Db 1420 G-----ACCGGAGAGAGCCCTGACCAAGCCTGAAGAGAGCTGTTCCGCAACGAGCCG 1467  
QY 1498 CTGAGCCAGTAA 1509  
Db 1468 CTGAGCCAGTAA 1479

RESULT 10  
AX468547  
LOCUS AX468547 1479 bp DNA linear PAT 16-JUL-2002  
DEFINITION Sequence 67 from Patent WO0226209.  
ACCESSION AX468547  
VERSION AX468547.1 GI:21901377  
KEYWORDS  
SOURCE Human immunodeficiency virus 1 (HIV-1)  
ORGANISM Human immunodeficiency virus 1  
REFERENCE 1  
AUTHORS O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singer,M., Ulmer,J. and Dubensky,T.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 67 04-APR-2002;  
CHIRON CORPORATION (US)  
FEATURES  
Source 1..1479  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"

ORIGIN  
Query Match 84.2%; Score 1271.2; DB 6; Length 1479;  
Best Local Similarity 91.8%; Pred. No. 3.6e-125;  
Matches 138; Conservative 10; Mismatches 78; Indels 36; Gaps 4;

QY 1 ATGGGCGCCCGCGGCAAGCATCTGCGCGGCGAGAGAGCTGACAACTGGAGAGATCCGC 60  
Db 1 ATGGGCGCCCGCGGCAAGCATCTGCGCGGCGAGAGAGCTGACCGCTGGAGCGCATCCGC 60  
QY 61 CTGGCGCCCGCGGCAAGAGCACTATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
Db 61 CTGGCGCCCGCGGCAAGAGTGTATGATGAAGCACTGTGTGGCCAGCCGCGAG 120  
QY 121 CTGAGAGGCTTGGCCCTGAACCCCGGCTGTGAGAGCCGCGAGGCTGCAAGAGATC 180  
Db 121 CTGAGAGAGTTCGCCCTGAACCCCGGCTGTGAGAGCCGAGGCTGCAAGAGATC 180  
QY 181 ATGAAGCAGCTGAGAGCCCGCTGAGAGCCGCGAGAGAGCTGCGCAGCTGTACAAC 240  
Db 181 ATCGGCCAGCTGAGAGCCCGCTGAGAGCCGCGAGAGAGCTGAAGAGCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGAGAGCCCGGCAATGAGGTCCGCGACACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTACTGCGTGAGAGCCCGGCAATGAGGTSCGCGACACCAAGAGGCC 300  
QY 301 CTGAGCAAGATCGAGAGAGAGCAAGTCCAGCAGAGAGCCAGAGGCCAAGAGAG 360  
Db 301 CTGAGCAAGATCGAGAGAGAGCAAGTCCAGCAGAGAGATCCAGAGGCCGAGGCC 360  
QY 361 GCGCAGC---GGCAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGAGGGCCAGATG 417  
Db 361 GCGCAGCAGGGCAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGAGGGCCAGATG 420  
QY 418 GTGACACGAGGCGCATGAGCCCGCGCACCTGAAGCCTGTGAGAGGTGATGAGAGAG 477

Db 421 GTGACACGAGGCCATCAGCCCCCGCACCTGAAGCGCTGGGTGAAGGTGATCGAGAGAG 480  
QY 478 GCGTTTACGCCCGAGAGTGTATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCAG 537  
Db 481 GCGTTTACGCCCGAGAGTGTATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
QY 538 GACCTGAACACGATGTGAAACACCGTGGGGGCCACAGGCCCGCATGACAGATGCTGAAG 597  
Db 541 GACTGAACACGATGTGAAACACCGTGGGGGCCACAGGCCCGCATGACAGATGCTGAAG 600  
QY 598 GACACCATCAACGAGAGAGGCCCGCGAGTGGAGCCGCTGACACCCCGTGAAGCCGCGCC 657  
Db 601 GACACCATCAACGAGAGAGGCCCGCGAGTGGAGCCGCTGACACCCCGTGAAGCCGCGCC 660  
QY 658 GTGGCCCCCGGAGATGCGCGACCCCGCGGAGCGACATCGCCGCGCCACAGCACC 717  
Db 661 ATGCGCCCCCGGAGATGCGCGAGCCCGCGGAGCGACATCGCCGCGCCACAGCACC 720  
QY 718 CTGCAAGAGCAGATCGCCTGTGATGACCAAGCAACCCCGGTGCGCGTGGGCGACATCTAC 777  
Db 721 CTGCAAGAGCAGATCGCCTGTGATGACCAAGCAACCCCGCATCCCGGTGGCGCATCTAC 780  
QY 778 AAGCGGTGATCATCTTGGGCTTGAACAGATGCTGGGATGTACAGCCCGTGAAGCATC 837  
Db 781 AAGCGGTGATCATCTTGGGCTTGAACAGATGCTGGGATGTACAGCCCGTGAAGCATC 840  
QY 838 CTGGAATATCCCGCAGAGGGCCCGCAAGAGCCCTTCCGCACTAGTGAAGCCGCTTCTTAAG 897  
Db 841 CTGGAATATCAGCAGAGGGCCCGCAAGAGCCCTTCCGCACTAGTGAAGCCGCTTCTTAAG 900  
QY 898 ACCCTGCGCGCGGAGCAGAGGCCACCGAGAGCTGAAGAACTGATGACCGAGACCTGCTG 957  
Db 901 ACCCTGCGCGCGGAGCAGAGGCCACCGAGAGCTGAAGAACTGATGACCGAGACCTGCTG 960  
QY 958 GTGCAAGACGCGCAACCCCGATGCAAGACCATCTGCGCGCTCTCGGCCCGCGCCACC 1017  
Db 961 GTGCAAGACGCGCAACCCCGATGCAAGACCATCTGCGCGCTCTCGGCCCGCGCCACC 1020  
QY 1018 CTGAGAGAGATGATGACCGCTGCGCAAGGCGGTGGGCGCCCGCCACAGAGCCCGCGTG 1077  
Db 1021 CTGAGAGAGATGATGACCGCTGCGCAAGGCGGTGGGCGCCCGCCACAGAGCCCGCGTG 1080  
QY 1078 CTGGCGGAGGCGATGAGCGCAGGCCCAAGCGGTGAACATCATGATGCAAGAGCAACTTC 1137  
Db 1081 CTGGCGGAGGCGSATGAGCGCAGGCCCAACAC---AGCGTATGATGCAAGAGCAACTTC 1137  
QY 1138 AAGGCGCCCGCGGCAAGCTGAAGTGTCTCAACTGCGGCAAGAGAGGCCCATGCGCCAG 1197  
Db 1138 AAGGCGCCCGCGGCAAGCTGAAGTGTCTCAACTGCGGCAAGAGAGGCCCATGCGCCAG 1197  
QY 1198 AACTGCGCGCGCGCGCAAGAGGCGCTGTGGAAGTGGCGCAAGAGAGGCCACAGATG 1257  
Db 1198 AACTGCGCGCGCGCGCGCAAGAGGCGCTGTGGAAGTGGCGCAAGAGAGGCCACAGATG 1257  
QY 1258 AAGGACTGCAACGAGCGCGCAAGCTTCTTGGGCAAGATCTGGCCCAAGCCCAAGGGC 1317  
Db 1258 AAGGACTGCAACGAGCGCGCAAGCTTCTTGGGCAAGATCTGGCCCAAGCCCAAGGGC 1317  
QY 1318 CGCCCCGGCACTTCTTGAAGAACCGCAGAGAGCCCGCGCCCAAGCTGCGCACCGCC 1377  
Db 1318 CGCCCCGGCACTTCTTGAAGAGCGG-----CCCGAGCCCAAGCCGCGC 1359  
QY 1378 CCCCCCGGAGAGCTTCGGCTTGAGAGAGACACCCCGCCCCCAAGAGAGGCCCAAG 1437  
Db 1360 CCCCCCGGAGAGCTTCGGCTTGAGAGAGACACCCCGCCCCAAGAGAGAGAGCAAG 1419  
QY 1438 GACCGGAGAGCCCTACCGGAGAGCCCTGACCGCGCTGCGAGAGCTGTTCCGAGCGGCC 1497  
Db 1420 G-----ACCGGAGAGAGCCCTGACCAAGCCTGAAGAGCTGTTCCGCAACGAGCCG 1467  
QY 1498 CTGAGCCAGTAA 1509

```

DB              1468 CTGAGCCAGTAA 1479

RESULT 11
AX455904
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
    source
        AX455904
        Sequence 20 from Patent WO0204493.
        AX455904
        AX455904.1  GI:21714896
        synthetic construct
        synthetic construct
        other sequences; artificial sequences.
1
zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,B.
Polynucleotides encoding antigenic hiv type c polypeptides,
Patent: WO 0204493-A 20 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
    location/Qualifiers
        1. 1479
        /organism="synthetic construct"
        /mol_type="unassigned DNA"
        /db_xref="taxon:32630"
        /note="synthetic Gag coding sequence of HIV strain
AF110965"

```

**ORIGIN**

Query Match	83.6%;	Score 1260.8;	DB 6;	Length 1479;
Best Local Similarity	91.9%;	Pred. No. 4.4e-124;		
Matches 1389; Conservative	0;	Mismatches 87;	Indels 36;	Gaps 4;

Qy	1	ATGGGCGCCCGCGCCAGCATCTGCGCGCGGAGAGCTGGACAACTGGGAGAAAGATCCGC	60
Db	1	ATGGGCGCCCGCGCCAGCATCTGCGCGCGGAGAGCTGGACAACTGGGAGCGCATCCGC	60
Qy	61	CTGCGCCCCCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCCAGCCGCCAG	120
Db	61	CTGCGCCCCCGCGGCAAGAGTGCTACATGATGAAGCACTGTGTGTGGGCCAGCCGCCAG	120
Qy	121	CTGGAGGGGCTTCGCCCCCTGAACCCCGGCTCTGTGAAGACCCCGAGGGCTGCAAGCAGATC	180
Db	121	CTGGAGGAAGTTCGCCCCCTGAACCCCGGCTCTGTGAAGACCGAGGGCTGCAAGCAGATC	180
Qy	181	ATGAAGCAGCTGCAGCCCCCGCTGCAGACCCGGCACCGGAGAGCTGCGCACCCTGTACAC	240
Db	181	ATCGGCCAGCTGCACCCCGCTGCAGACCCGGCAGCGAGGAGCTGAAGAGCCTGTTCAAC	240
Qy	241	ACCGTGGCCACCCCTGTACTGCGTGCAAGCCGGCATCGAGGTCGCGGACCAAGAGAGGC	300
Db	241	ACCGTGGCCACCCCTGTACTGCGTGCAAGAGATCGAGGTGCGCGGACCAAGAGAGGC	300
Qy	301	CTGGAACAAGATCGAGGAGGAGCAGAACAAGTCCCAAGCAGAAACCAGCAGGCCAAGAGAG	360
Db	301	CTGGAACAAGATCGAGGAGGAGCAGAACAAGTCCCAAGCAGAAATCCAGCAGGCCGAGGCC	360
Qy	361	GCCGAC--GGCAAAGGTAGCCAGAACTACCCCATCGTGCAAACTGCAAGGGCCAGATG	417
Db	361	GCCGACAAAGGCGCAAAGGTAGCCAGAACTACCCCATCGTGCAAACTGCAAGGGCCAGATG	420
Qy	418	GTGCAACAGGCCCATCAGCCCCCGCACCTGAACGCGCTGGGTGAAGGTGATCGAGGAGAG	477
Db	421	GTGCAACAGGCCCATCAGCCCCCGCACCTGAACGCGCTGGGTGAAGGTGATCGAGGAGAG	480
Qy	478	GCCTTACGCCCCCGAGGTGATCCCATGTTCAACCGGCTTGAGCGGAGGGGCCACCCCCCAG	537
Db	481	GCCTTACGCCCCCGAGGTGATCCCATGTTCAACCGGCTTGAGCGGAGGGGCCACCCCCCAG	540
Qy	538	GACCTGAACAACGATGTTGAACAACCGTGGGGCGGCCAACAGGCCCATGCAATGCTGTGAAG	597
Db	541	GACCTGAACAACCATGCTGAACAACCGTGGGGCGGCCAACAGGCCCATGCAATGCTGTGAAG	600
Qy	598	GACACCATCAACGAGGAGGCGCGCGAGTGGGACCGGCTTGCAACCCCGTGCAAGCGCGGCC	657

Db	601	GACACCATCAACGAGAGGCCCGAGTGGGACCGGTGACCCCGTGACCGCGGCC	660
QY	658	GTGGCCCCCGGCAGATGCGCGCAACCCCGCGGACGCAATCGCCGGCCACGACACC	717
Db	661	ATCGCCCCCGGCAGATGCGCGAGCCCGCGGACGCAATCGCCGACCAACGACACC	720
QY	718	CTGCAGAGCAGATCGCTGGATGACCAACCCCGGTGCCGTGGCGCATCTAC	777
Db	721	CTGCAGAGCAGATCGCTGGATGACCAACCCCGCATCCCGTGGCGCATCTAC	780
QY	778	AAGCGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGTGAGCATC	837
Db	781	AAGCGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGTGAGCATC	840
QY	838	CTGCATATCCGCCAGGAGCCCAAGAGCCCTCCGCACTACGTGACCGCTTCTTCAAG	897
Db	841	CTGCATATCAAGCAGGAGCCCAAGAGCCCTCCGCACTACGTGACCGCTTCTTCAAG	900
QY	898	ACCCCTGCGCGGAGCAGGCGCACCGAGAGCTGAAGACTGATGACCGAGACCTGCTG	957
Db	901	ACCCCTGCGCGGAGCAGGCGCACCGAGAGCTGAAGACTGATGACCGAGACCTGCTG	960
QY	958	GTGCAGAACGCCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGCCCGGCGCAC	1017
Db	961	GTGCAGAACGCCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGCCCGGCGCAC	1020
QY	1018	CTGAGAGATGATGACCGCTGCGAGGCGTGGGCGCCCGGCGCAAGAGCCCGGTG	1077
Db	1021	CTGAGAGATGATGACCGCTGCGAGGCGTGGGCGCCCGGCGCAAGAGCCCGGTG	1080
QY	1078	CTGCCCGAGGCGGATGAGCCAGGCCCAACAGCGTGAATCATGATGAGAGCACTTC	1137
Db	1081	CTGCCCGAGGCGGATGAGCCAGGCCCAACAC---AGCGTATGATGAGAGCACTTC	1137
QY	1138	AAGGGCCCCCGGCGCAACGTCATGCTTCAACTGCGGCAAGAGGGCGCATCGCCAG	1197
Db	1138	AAGGGCCCCCGGCGCATCGTGAAGTGCTTCAACTGCGGCAAGAGGGCGCATCGCCG	1197
QY	1198	AACTGCGCGCCCCCGGCAAGAGGCTGCTGGAAGTGCAGCAAGAGGGCCACAGATG	1257
Db	1198	AACTGCGCGCCCCCGGCAAGAGGCTGCTGGAAGTGCAGCAAGAGGGCCACAGATG	1257
QY	1258	AAGACTGCAACCGAGCGCCAGGCCCACTTCTTGCGCAAGATCTGGCCAGCCACAAGGG	1317
Db	1258	AAGACTGCAACCGAGCGCCAGGCCCACTTCTTGCGCAAGATCTGGCCAGCCACAAGGG	1317
QY	1318	CGCCCCGCGCACTTCTGCGAAGACCGCAGCGAGCCCGCGCCCGCCACCGTGCACCGC	1377
Db	1318	CGCCCCGCGCACTTCTGCGAAGACCGCAGCGAGCCCGCGCGCCCGCCACCGTGCACCGC	1377
QY	1378	CCCCCGCGGAGAGCTTCGCTTGAAGAGACCAACCCCGCCCGCCCAAGCAGAGCCCAAG	1437
Db	1360	CCCCCGCGGAGAGCTTCGCTTGAAGAGACCAACCCCGCGCCCAAGCAGAGCAAG	1419
QY	1438	GACCGGAGGCTTACCGCGAGCCCTGTACCGCCCTGCGGAGCTGTCGCGAGCGGCCCC	1497
Db	1420	G-----ACCGGAGACCTGTACCGCTGAAGAGCTGTCGCGAGCGGCCCC	1467
QY	1498	CTGAGCCAGTAA 1509	
Db	1468	CTGAGCCAGTAA 1479	

RESULT 12			
AX149648			
LOCUS	AX149648	4288 bp	DNA
DEFINITION	Sequence 2 from Patent WO0136614.		linear
ACCESSION	AX149648		PAT 08-JUN-2001
VERSION	AX149648.1	GI:14348047	
KEYWORDS			
SOURCE	Human immunodeficiency virus		
ORGANISM	Human immunodeficiency virus		

Viruses; Retrovird viruses; Retroviridae; Lentivirus; Primate  
lentivirus group.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1  
Shao,Y., Wagner,R., Wolf,H. and Graf,M.  
The genome of the hiv-1 inter-subtype (c/b') and use thereof  
Patent: WO 0136614-A 2 25-MAY-2001;  
Geneart GMBH Gesellschaft fuer angewandte Biotechnologie (DE) ;  
Shao, Yiming (CN)

FEATURES  
source  
1..4288  
Location/Qualifiers  
/organism="Human immunodeficiency virus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:12721"

ORIGIN

Query Match 79.7%; Score 1202.2; DB 6; Length 4288;  
Best Local Similarity 88.5%; Pred. No. 4.9e-118;  
Matches 1335; Conservative 0; Mismatches 153; Indels 21; Gaps 2;

QY 1 ATGGGCGCCGCGCCAGCATCTCGCGCGGAGAGCTGACAACTGGGAGAATCCGC 60  
DB 13 ATGGGCGCCAGGCGCAGCATCTGAGGGGCGGCAAGCTGACAACTGGGAGAATCAGG 72  
QY 61 CTGCGCCCGCGCGGCAAGACATGCTGAGCACTGTGTGGGCCAGCCCGGAG 120  
DB 73 CTGAGGCGCCGCGGCAAGACATGCTGAGCACTGTGTGGGCCAGCGGAG 132  
QY 121 CTGAGGGCTTTCGCTTGAACCCCGCTGTGAGACCGCGGCTGCAAGCATC 180  
DB 133 CTGAGAGGTTTCGCTTGAACCCCGCTGTGAGACCAAGCGGCTGCAAGCATC 192  
QY 181 ATGAAGCAGCTGACGCGCGCTGAGACCGGACCGAGAGCTGCGCAGCTTGAAC 240  
DB 193 ATGAAGCAGCTGACGCGCGCTGAGACCGGACCGAGAGCTGCGCAGCTTGAAC 252  
QY 241 ACCGTGGCCACCTGTACTGTGTCGACGCGCGCATCGAGTCCGCGACCAAGAGGCC 300  
DB 253 ACCGTGGCCACCTGTACTGTGTCGACGCGCGCATCGAGTCCGCGACCAAGAGGCC 312  
QY 301 CTGGAACAAGATCGAGAGGAGCAGAACATGCCAGACAGAACCCAGCAGGCCAAGAG 360  
DB 313 CTGGAACAAGATCGAGAGGAGCAGAACATGCCAGACAGAACCCAGCAGGCCAAGAG 372  
QY 361 GCGGACGGCAAGGTGAGCCAGAACTACCCCATCTGTCAGAACCTGCAAGGCCAGATG 420  
DB 373 GCGGACGGCAAGGTGAGCCAGAACTACCCCATCTGTCAGAACCTGCAAGGCCAGATG 432  
QY 421 CACCAAGGCGCATCAGCCCCCGCAGACCTGAAACGCTGGGTGAAGTGCATCGAGAGAGGCC 480  
DB 433 CACCAAGGCGCATCAGCCCCCGCAGACCTGAAATGCATGGGTGAAGTGTGAGAGAGAGGCC 492  
QY 481 TTCAGCCCCGAGGTGATCCCATGTTTACCGGCTGAGCGAGGGGCCACCCCCAGAGAC 540  
DB 493 TTCAGCCCCGAGGTGATCCCATGTTTACCGGCTGAGCGAGGGGCCACCCCCAGAGAC 552  
QY 541 CTGAACAAGATGTTGAACACCGTGGGGCGGCCACAGGCCGCGCATGCAAGTGTGAAGAG 600  
DB 553 CTGAACAAGATGTTGAACACCGTGGGGCGGCCACAGGCCGCGCATGCAAGTGTGAAGAG 612  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGGACCGCTGCACCCCGTGCAGGCCGCGCCGCTG 660  
DB 613 ACCATCAACGAGAGGCGCGCGAGTGGGACAGGCTGCACCCCGTGCAGGCCGCGCCGCTG 672  
QY 661 GCGCCCGGCGCAGATGCGGACCGCGCGCGCAGCAGCATCGCGCGGCCACCAAGACCCCTG 720  
DB 673 GCGCCCGGCGCAGATGCGGACCGCGCGCGCAGCAGCATCGCGCGGCCACCAAGACCCCTG 732  
QY 721 CAGGAGCAGATCGCTGTGATGACCAAGAACCCCGCTGCGCGTGGGGCGCATCTACAAG 780  
DB 733 CAGGAGCAGATCGCTGTGATGACCAAGAACCCCGCTGCGCGTGGGGCGCATCTACAAG 792  
QY 781 CGGTGATCATCTGTGGCTGAAACAAGATCGTGGGATGTACAGCCCGCTGAGCATCTG 840

DB 793 AGGTGATCATCTGGGTTTAAACAAGATCGTGAAGATGTACAGCCCAACCAAGCATCTG 852  
QY 841 GACATCCGCGCAGGCGCCCAAGAGACCTTCCGCGACTACGTGAGACCGCTTCTCAAGAC 900  
DB 853 GACATCAAGCAGGCGCCCAAGAGACCTTCAAGGACTACGTGACAGGTTCTTCAAGAC 912  
QY 901 CTGCGCGCGCAGGCGCCCAAGAGAGCTGAAAGAACTGATGACCGAGACCTGTGCTG 960  
DB 913 CTGAGGGCGGAGCAGGCGCACCAAGGCGGTGAAGAACTGATGACCGACACCTGTGCTG 972  
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCGCGGCCACCTG 1020  
DB 973 CAGAACGCCAACCCCGACTGCAAGACCATCTGAGGGCGCTGCGCCCGCGGCCAGCATC 1032  
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGGTGCGCGCGCCCGGCCCAAGGCGCGTGTG 1080  
DB 1033 GAGGAGATGATGACCGCTGCGCAGGCGGTGCGCGCGCCCGGCCCAAGGCGCGTGTG 1092  
QY 1081 GCGGAGCGATGACCGCGCCCAAGCGGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
DB 1093 GCGGAGCGCATGACCGCGCCCAAGCGG---CATCTGATGCAAGAGCAACTTCAAG 1149  
QY 1141 GCGCCCGCGCGCAAGCTCAAGTGTCTTCAACTGCGGCAAGAGGCGCCACATGCGCAAGAC 1200  
DB 1150 GCGCAACAAGAGATCGTGAAGTGTCTTCAACTGCGGCAAGAGGCGCCACATGCGCAAGAC 1209  
QY 1201 TGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGCAGGCAAGAGGCGCCACAGATGAAG 1260  
DB 1210 TGCAAGGCGCGCAAGAGAGGCGTGTGAAAGTGCAGGCAAGAGGCGCCACAGATGAAG 1269  
QY 1261 GACTGACCGAGCGCGCGCAAGCTTCTGCGCAAGATGTGCGCCAGCAAGAGGCGCG 1320  
DB 1270 GACTGACCGAGAGGCGCGCAAGCTTCTGCGCAAGATGTGCGCCAGCAAGAGGCGCG 1329  
QY 1321 CCGGCAACTTCTGCAAGACCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
DB 1330 CCGGCAACTTCTGCAAGACG-----GCGGAGCGCGCGCGCGCGCGCGCGCGCG 1371  
QY 1381 CCGGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCAAGAGAGCGCGCAAGAG 1440  
DB 1372 CCGGAGAGAGCTTCAAGTTCAGAGAGAGACCAACCCCGCGCGCAAGAGAGAGAGCGCG 1431  
QY 1441 CCGGAGCGCTTACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
DB 1432 ATGCAACAAGAGCTGTACCCCTTGACCAAGCGCTGTAAGAGCGCTGTTCGCAAGACCCAGC 1491  
QY 1501 AGCCAGTAA 1509  
DB 1492 AGCCAGTAA 1500

RESULT 13  
AF201927  
LOCUS 1509 bp DNA linear SYN 16-MAR-2000  
DEFINITION Synthetic construct gag protein gene, complete cds.  
ACCESSION AF201927  
VERSION AF201927.1 GI:7248702  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
other sequences; artificial sequences.  
REFERENCE  
1 (bases 1 to 1509)  
AUTHORS  
zur Megede,J., Chen,M.C., Doe,B., Schaefer,M., Greer,C.B.,  
Selby,M., Otten,G.R. and Barnett,S.W.  
TITLE  
Increased expression and immunogenicity of sequence-modified human  
immunodeficiency virus type 1 gag gene  
JOURNAL  
J. Virol. 74 (6), 2628-2635 (2000)  
MEDLINE  
20148954  
PUBMED  
10684277  
REFERENCE  
2 (bases 1 to 1509)  
AUTHORS  
zur Megede,J. and Barnett,S.W.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (04-NOV-1999) Vaccines, Chiron Corporation, 4560 Horton,

Emeryville, CA 94608, USA  
Location/Qualifiers

FEATURES

Source

CDS

1..1509  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
1..1509  
/note="HIV-1SF2 p55gag"  
/codon\_start=1  
/transl\_table=11  
/product="gag protein"  
/protein\_id="AAFA3628.1"  
/db\_xref="GI:7248703"  
/translation="MGARASVLSGELDKWEKIRLRPGKKKKYKIKHIVASRELERF  
AVNPGLTSEGCRIQLGQLPSLQTSSEELSLYNTVATLYCVHQRIDKDTKEALE  
KIEBQNKSKKAQAAAGAGTGNSSQVSNYPIVQNLQGMVHQALISPTLNAMVKV  
VEKAFSPVPIPMFSALSGATPDNLNTMLTVGHOAQMOKETINSEBAEMDRVH  
PVHAGPIAPGQMBREPRGSDIAGTSTLQEOIGMTNNPPIPVGEIYKRWIILGLNKIV  
RMYSPISLIDIRQGPKEPFIDYVDFYKTLRAEQASQDVKNMTETLLVQNPANPDCT  
ILKALGPATLBEEMTACQGVGPGHKARVLAEMASQVTPNATIMQGRNFRNQKTV  
KCFNCGKEGHTARNCRAPRKGCWRCRGEGHQMDCTERQANFLGKIWPYKGRPGNF  
LQSRPEPTAPBESFRFGEKRTTPSQKQEPIDKELYPLTSLRSLFGNDPSQ"

ORIGIN

Query Match 79.6%; Score 1201; DB 12; Length 1509;  
Best Local Similarity 88.5%; Pred. No. 8.8e-118;  
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;

QY 1 ATGGGCGCCGCGCCAGCATCTCGCGCGGAGAGAGTGAACAAGTGGAGAGAATCCGC 60  
DB 1 ATGGGCGCCGCGCGCCAGCATCTCGCGCGGAGAGAGTGAACAAGTGGAGAGAATCCGC 60  
QY 61 CTGCGCCCGCGCGGCAAGAAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
DB 61 CTGCGCCCGCGCGGCAAGAAGCACTACATGCTGTGTGGCCAGCCGCGAG 120  
QY 121 CTGAGGGCTTTCGCTGAAACCCGCGCTGTGAAGACCGCCGAGGCTGCAAGCAGATC 180  
DB 121 CTGAGGGCTTTCGCTGAAACCCGCGCTGTGAAGACCGCCGAGGCTGCAAGCAGATC 180  
QY 121 CTGAGCGCTTTCGCTGAAACCCGCGCTGTGAAGACCGCCGAGGCTGCAAGCAGATC 180  
DB 121 CTGAGCGCTTTCGCTGAAACCCGCGCTGTGAAGACCGCCGAGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGACGCGCCCTGCAAGCCGCAAGAGAGTGCAGCAGCTGTACAC 240  
DB 181 ATGAAGCAGCTGACGCGCCCTGCAAGCCGCAAGAGAGTGCAGCAGCTGTACAC 240  
QY 181 CTGGGCGAGCTGACGCGCCCTGCAAGCCGCAAGAGAGTGCAGCAGCTGTACAC 240  
DB 181 CTGGGCGAGCTGACGCGCCCTGCAAGCCGCAAGAGAGTGCAGCAGCTGTACAC 240  
QY 241 ACCGTGGCCACCTGTACTGCTGACAGCCGCGCATGAGTCCGCAACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGCTGACAGCCGCGCATGAGTCCGCAACCAAGAGGCC 300  
QY 241 ACCGTGGCCACCTGTACTGCTGACAGCCGCGCATGAGTCCGCAACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGCTGACAGCCGCGCATGAGTCCGCAACCAAGAGGCC 300  
QY 301 CTGACCAAGATCGAGAGGAGCAGAACAGTCCAGAGAGAGGCCGCGCGCC 360  
DB 301 CTGACCAAGATCGAGAGGAGCAGAACAGTCCAGAGAGAGGCCGCGCGCC 360  
QY 361 GCCGACGGCA-----AGGTGAGCCAGAACTACCCCATCTGTGCAAGACTG 405  
DB 361 GCCGACGGCA-----AGGTGAGCCAGAACTACCCCATCTGTGCAAGACTG 405  
QY 361 GCCGCGCGGCAACGAGCAGCCAGTGAAGCAACTACCCCATCTGTGCAAGACTG 420  
DB 361 GCCGCGCGGCAACGAGCAGCCAGTGAAGCAACTACCCCATCTGTGCAAGACTG 420  
QY 406 CAGGCGCAGATGCTGACCAAGGCGCATGAGCCCGCAACCTGAACGCTGGTGAAGTG 465  
DB 406 CAGGCGCAGATGCTGACCAAGGCGCATGAGCCCGCAACCTGAACGCTGGTGAAGTG 465  
QY 421 CAGGCGCAGATGCTGACCAAGGCGCATGAGCCCGCAACCTGAACGCTGGTGAAGTG 480  
DB 421 CAGGCGCAGATGCTGACCAAGGCGCATGAGCCCGCAACCTGAACGCTGGTGAAGTG 480  
QY 466 ATCGAGAGAGAGGCTTTCAGCCCGGAGGTGATCCCATGTTTACCCGCTGAGCGAGGC 525  
DB 466 ATCGAGAGAGAGGCTTTCAGCCCGGAGGTGATCCCATGTTTACCCGCTGAGCGAGGC 525  
QY 481 GTGAGAGAGAGGCTTTCAGCCCGGAGGTGATCCCATGTTTACCGCTGAGCGAGGC 540  
DB 481 GTGAGAGAGAGGCTTTCAGCCCGGAGGTGATCCCATGTTTACCGCTGAGCGAGGC 540  
QY 526 GCCACCCCCCAGAGACTGAACAGATTTGAACACCGTGGGGCGGCAACAGGCGCGCATG 585  
DB 526 GCCACCCCCCAGAGACTGAACAGATTTGAACACCGTGGGGCGGCAACAGGCGCGCATG 585  
QY 541 GCCACCCCCCAGAGACTGAACAGATTTGAACACCGTGGGGCGGCAACAGGCGCGCATG 600  
DB 541 GCCACCCCCCAGAGACTGAACAGATTTGAACACCGTGGGGCGGCAACAGGCGCGCATG 600  
QY 586 CAGATGCTGAAGAGACCATCAACGAGAGAGGCGCGGAGTGGGACCGGCTGCACCCCGTG 645  
DB 586 CAGATGCTGAAGAGACCATCAACGAGAGAGGCGCGGAGTGGGACCGGCTGCACCCCGTG 645  
QY 601 CAGATGCTGAAGAGACCATCAACGAGAGAGGCGCGGAGTGGGACCGGCTGCACCCCGTG 660  
DB 601 CAGATGCTGAAGAGACCATCAACGAGAGAGGCGCGGAGTGGGACCGGCTGCACCCCGTG 660  
QY 646 CAGGCGGCGCCGCTGGCGCCCGGCAAGATGCGCGACCCCGCGGCAAGCAGATCGCGCGC 705  
DB 646 CAGGCGGCGCCGCTGGCGCCCGGCAAGATGCGCGACCCCGCGGCAAGCAGATCGCGCGC 705

DB 661 CAGCGCGCGCCCATCGCCCCCGGCGAGATGCGCGAGCCCCCGGCGAGCAGATCGCGCGC 720  
QY 706 GCCACCGAGCAGCCCTGAGAGAGAGATCGCTGATGACCAAGCAACCCCGTGGCGGTG 765  
DB 706 GCCACCGAGCAGCCCTGAGAGAGAGATCGCTGATGACCAAGCAACCCCGTGGCGGTG 765  
QY 721 ACCACCGAGCAGCCCTGAGAGAGAGATCGCTGATGACCAAGCAACCCCGTGGCGGTG 780  
DB 721 ACCACCGAGCAGCCCTGAGAGAGAGATCGCTGATGACCAAGCAACCCCGTGGCGGTG 780  
QY 766 GGGCAGCATCTAAGCGGTGATGATCTGAGGCTGAACAAGATGCTGGATGACAGC 825  
DB 766 GGGCAGCATCTAAGCGGTGATGATCTGAGGCTGAACAAGATGCTGGATGACAGC 825  
QY 781 GGGCAGATCTAAGCGGTGATGATCTGAGGCTGAACAAGATGCTGGATGACAGC 840  
DB 781 GGGCAGATCTAAGCGGTGATGATCTGAGGCTGAACAAGATGCTGGATGACAGC 840  
QY 826 CCGGTGAGCATCTGAGCATCCGCGAGGCGCCCAAGAGGCCCTTCGCGACTACGTGAGC 885  
DB 826 CCGGTGAGCATCTGAGCATCCGCGAGGCGCCCAAGAGGCCCTTCGCGACTACGTGAGC 885  
QY 841 CCGACAGCATCTGAGCATCCGCGAGGCGCCCAAGAGGCCCTTCGCGACTACGTGAGC 900  
DB 841 CCGACAGCATCTGAGCATCCGCGAGGCGCCCAAGAGGCCCTTCGCGACTACGTGAGC 900  
QY 886 CGCTTCTTCAAGACCTGCGCGCGAGAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGC 945  
DB 886 CGCTTCTTCAAGACCTGCGCGCGAGAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGC 945  
QY 901 CGCTTCTAAGAGACCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 960  
DB 901 CGCTTCTAAGAGACCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 960  
QY 946 GAGACCTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1005  
DB 946 GAGACCTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1005  
QY 961 GAGACCTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1020  
DB 961 GAGACCTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1020  
QY 1006 CCGCGCGCCACCTGAGAGAGAGATGATGACCGCTGCCAGGCGGTGGCGCGCGCGCAC 1065  
DB 1006 CCGCGCGCCACCTGAGAGAGAGATGATGACCGCTGCCAGGCGGTGGCGCGCGCAC 1065  
QY 1021 CCGCGCGCCACCTGAGAGAGATGATGACCGCTGCCAGGCGGTGGCGCGCGCAC 1080  
DB 1021 CCGCGCGCCACCTGAGAGAGATGATGACCGCTGCCAGGCGGTGGCGCGCGCAC 1080  
QY 1066 AAGCGCGCGGTGCTGGCGCGAGGCGATGAGCCAGG---CCAACAGCGTGAACATGATG 1122  
DB 1066 AAGCGCGCGGTGCTGGCGCGAGGCGATGAGCCAGG---CCAACAGCGTGAACATGATG 1122  
QY 1081 AAGCGCGCGGTGCTGGCGCGAGGCGATGAGCCAGGCGTGAACAGAGAGAGAGAGAGAG 1140  
DB 1081 AAGCGCGCGGTGCTGGCGCGAGGCGATGAGCCAGGCGTGAACAGAGAGAGAGAGAG 1140  
QY 1123 CAGAAAGCACTTCAAGAGGCGCGCGCGCGCAAGCTCAAGTCTTCAACTGCGGCAAGAG 1182  
DB 1123 CAGAAAGCACTTCAAGAGGCGCGCGCGCGCGCAAGCTCAAGTCTTCAACTGCGGCAAGAG 1182  
QY 1141 CAGCGCGCGCACTTCCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
DB 1141 CAGCGCGCGCACTTCCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 1183 GGGCAGCATGCGCAAGAACTGCGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAG 1242  
DB 1183 GGGCAGCATGCGCAAGAACTGCGCGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAG 1242  
QY 1201 GGGCAGCATGCGCAAGAACTGCGCGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAG 1260  
DB 1201 GGGCAGCATGCGCAAGAACTGCGCGCGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAG 1260  
QY 1243 GAGGCGCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302  
DB 1243 GAGGCGCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302  
QY 1261 GAGGCGCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1261 GAGGCGCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
QY 1303 CCGAGCCCAAGAGGCGCGCGCGCGCACTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAG 1362  
DB 1303 CCGAGCCCAAGAGGCGCGCGCGCGCGCACTTCTGCAAGAGAGAGAGAGAGAGAGAGAG 1362  
QY 1321 CCGAGCTCAAGAGGCGCGCGCGCGCACTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAG 1382  
DB 1321 CCGAGCTCAAGAGGCGCGCGCGCGCGCACTTCTGCAAGAGAGAGAGAGAGAGAGAGAG 1382  
QY 1363 ACCGTGGCCACCGCGCGCGCGCGCGAGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAG 1422  
DB 1363 ACCGTGGCCACCGCGCGCGCGCGCGCGAGAGAGCTTCCGCTTGAAGAGAGAGAGAGAG 1422  
QY 1363 CCGAGCGCCCAAGAGGCGCGCGCGCGCGAGAGAGCTTCCGCTTGAAGAGAGAGAGAGAG 1442  
DB 1363 CCGAGCGCCCAAGAGGCGCGCGCGCGCGCGAGAGAGCTTCCGCTTGAAGAGAGAGAGAG 1442  
QY 1423 AAGCAGAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
DB 1423 AAGCAGAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
QY 1483 TTGCGCAGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1509  
DB 1483 TTGCGCAGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1509

RESULT 14  
BD263637 1515 bp DNA linear PAT 17-JUL-2003  
LOCUS Improved expression of HIV polypeptides and production of  
DEFINITION virus-like particles.  
ACCESSION BD263637  
VERSION BD263637.1 GI:33073405  
KEYWORDS JP 2002533124-A/4.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 1515)  
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,

TITLE Greer,C., Selby,M. and Walker,C.  
Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: JP 200253124-A 4 08-OCT-2002;

COMMENT CHIRON CORP  
OS Artificial Sequence  
PN JP 200253124-A/4  
PD 08-OCT-2002  
PF 30-DEC-1999 JP 2000591193  
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI  
SUSAN BARNETT,JAN ZUR MEGBEDE,INDRESH SRIVASTAVA,YING LIAN, PI  
KARIN HARTOG,  
PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC  
C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC  
C12N5/10,  
PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC  
Description of Artificial Sequence: synthetic HIV-Gag FH Key

FT source 1..1515  
Location/Qualifiers  
1..1515  
/organism="Artificial Sequence".  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

FEATURES  
source  
location/Qualifiers  
1..1515  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

ORIGIN

Query Match 79.6%; Score 1201; DB 6; Length 1515;  
Best Local Similarity 88.5%; Pred. No. 8.8e-118;  
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;

QY 1 ATGGGCGCCCGCGCGCATCTCGCGCGCGGAGAGCTGGACAAGTGGAGAAATCCGC 60  
Db 7 ATGGGCGCCCGCGCGCATCTCGCGCGCGGAGAGCTGGACAAGTGGAGAAATCCGC 66  
QY 61 CTGGCGCCCGCGCGGCAAGAAGCACTAATCTGTAAGCACTGTGTGGGCGGCGGAG 120  
Db 67 CTGGCGCCCGCGCGGCAAGAAGCACTAATCTGTAAGCACTGTGTGGGCGGCGGAG 126  
QY 121 CTGAGAGGCTTCGCGCTGAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCATC 180  
Db 127 CTGAGAGGCTTCGCGCTGAACCCCGGCTGTGAGACCGCGGAGGCTGCGGAGATC 186  
QY 181 ATGAAGCAGCTGACGCGCGCGCTGACACCGGCAAGAGAGCTGCGGAGCTGTACAAC 240  
Db 187 CTGGGCGCAGCTGACGCGCGCGCTGACACCGGCAAGAGAGCTGCGGAGCTGTACAAC 246  
QY 241 ACCGTGGCCACCTGTACTGCTGTGACGCGCGCATTCGAGTCCGCGACACCAAGAGGCC 300  
Db 247 ACCGTGGCCACCTGTACTGCTGTGACGCGCGCATTCGAGTCCGCGACACCAAGAGGCC 306  
QY 301 CTGACAAGATCGAAGAGAGGACAAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 307 CTGAGAAGATCGAAGAGAGGAGCAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCCGACGGCA-----AGGTGAGCCAGAACTACCCCATCTGTGCAAACTG 405  
Db 367 GCCGCGCGGCAACCGGCAACAGCAGGAGTGAAGCAACTACCCCATCTGTGCAAACTG 426  
QY 406 CAGGGCCAGATGTGACCAAGGCGCATCAGCCCGGCAACCTGAAAGCCTGGGTGAAGTG 465  
Db 427 CAGGGCCAGATGTGACCAAGGCGCATCAGCCCGGCAACCTGAAAGCCTGGGTGAAGTG 486  
QY 466 ATCGAGAGAGAGGCTTCAGCCCGGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGC 525  
Db 487 GTGAGAGAGAGAGGCTTCAGCCCGGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGC 546  
QY 526 GCCACCCCGGAGAGAGCTGAAACAGATGTTGAACACCGTGGGCGGCGCAACGAGCGGCATG 585  
Db 547 GCCACCCCGGAGAGAGCTGAAACAGATGTTGAACACCGTGGGCGGCGCAACGAGCGGCATG 606  
QY 586 CAGATGCTGAAGAGACCATCAAGAGAGGCGCGCGAGTGGAGCCGCTGCAACCCCGTG 645

Db 607 CAGATGCTGAAGAGAGACCATCAAGAGAGGCGCGGAGTGGAGACCGGTGCAACCCCGTG 666  
QY 646 CAGCGCGCGCGGTGGCGCGCGCGCGAGATGCGCGGAGAGAGAGAGAGAGAGAGAGAG 705  
Db 667 CAGCGCGCGCGCGGTGGCGCGCGCGCGAGATGCGCGGAGAGAGAGAGAGAGAGAGAGAG 726  
QY 706 GCCACGAGAGAGAGAGAGAGAGATGCGGTGATGACCAAGAGAGAGAGAGAGAGAGAGAG 765  
Db 727 ACCACGAGAGAGAGAGAGAGAGATGCGGTGATGACCAAGAGAGAGAGAGAGAGAGAGAG 786  
QY 766 GCGGAGATCTCAAGCGGTGATCATCTGGGCTGAAACAGAGATGCGGAGTGTACAGC 825  
Db 787 GCGGAGATCTCAAGCGGTGATCATCTGGGCTGAAACAGAGATGCGGAGTGTACAGC 846  
QY 826 CCGGTGAGCATCTTGACATCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885  
Db 847 CCGACGAGCATCTTGACATCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906  
QY 886 CGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945  
Db 907 CGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966  
QY 946 GAGACCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005  
Db 967 GAGACCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026  
QY 1006 CCGGCGCGAGAGAGAGATGATGACCGCTGCCAGGCGGTGGCGCGCGCGCGAG 1065  
Db 1027 CCGGCGCGAGAGAGAGATGATGACCGCTGCCAGGCGGTGGCGCGCGCGCGAG 1086  
QY 1066 AAGCGCGGTGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1122  
Db 1087 AAGCGCGGTGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146  
QY 1123 CAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1182  
Db 1147 CAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206  
QY 1183 GGGCAGATGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1242  
Db 1207 GGGCAGATGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266  
QY 1243 GAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302  
Db 1267 GAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326  
QY 1303 CCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1362  
Db 1327 CCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368  
QY 1363 ACCGTGGCCACCGCGCGCGCGCGAGAGCTTCGCTGAGAGAGAGAGAGAGAGAGAGAG 1422  
Db 1369 CCCGAGCCACCGCGCGCGCGCGAGAGAGCTTCGCTGAGAGAGAGAGAGAGAGAGAGAG 1428  
QY 1423 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
Db 1429 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488  
QY 1483 TTCGGCAGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1509  
Db 1489 TTCGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1515

RESULT 15  
CQ870496 1515 bp DNA linear PAT 13-SEP-2004  
LOCUS CQ870496  
DEFINITION Sequence 4 from Patent EP1433851.  
ACCESSION CQ870496  
VERSION CQ870496.1 GI:52000023  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
other sequences; artificial sequences.

REFERENCE 1  
AUTHORS Barnett, S., Zurmehede, J., Sriwastava, I., Lian, Y., Hartog, K.,  
Liu, H., Greer, C., Selby, M. and Walker, C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: EP 1433851-A 4 30-JUN-2004;  
CHIRON CORPORATION (US)  
FEATURES Location/Qualifiers  
source 1..1515  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Description of Artificial Sequence: synthetic  
HIV-Gag"  
ORIGIN  
Query Match 79.6%; Score 1201; DB 6; Length 1515;  
Best Local Similarity 88.5%; Pred. No. 8.8e-118;  
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;  
QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGGAGAGCTGAGCAAGTGGAGAAATCCGC 60  
DB 7 ATGGGCGCGCGCGCAGCGTGTGAGCGCGGCGAGCTGAGCAAGTGGAGAAATCCGC 66  
QY 61 CTGCGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
DB 67 CTGCGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG 126  
QY 121 CTGAGGCGCTTCCGCTGAACCCCGGCTGTGAGAGACCGCGAGGCGTCAAGCAGATC 180  
DB 127 CTGAGGCGCTTCCGCTGAACCCCGGCTGTGAGAGACCGAGGCGTCCGCGCAGATC 186  
QY 181 ATGAAGCAGCTGACGCGCGCTGCAAGACCGGACCGAGAGCTGCGCAGCTGTACAAC 240  
DB 187 CTGCGCAGCTGACGCGCGCTGCAAGACCGGACCGAGAGCTGCGCAGCTGTACAAC 246  
QY 241 ACCGTGGCCACCTGTACTGCTGTCAGCGCGGATCGAGTCCGCGACCAAGAGGCC 300  
DB 247 ACCGTGGCCACCTGTACTGCTGTCAGCGCGGATCGAGTCCGCGACCAAGAGGCC 306  
QY 301 CTGACAAGATCGAGAGAGAGAGCAAGTCCAGCAGAAAGCCAGAGGCCAAGAG 360  
DB 307 CTGAGAAGATCGAGAGAGAGAGCAAGTCCAAAGAAAGGCCAGAGGCCGCGCC 366  
QY 361 GCCGACGGCA-----AGGTAGCCAGAACTACCCCATGTGTCAGAACTG 405  
DB 367 GCCGCGGGCACCGGCAACAGACGACGAGTGAAGCAAGAACTACCCCATGTGTCAGAACTG 426  
QY 406 CAGGCGCAGATGTGTCACAGGCGCATGACCCCGCACCTGAACGCTGGTGAAGTG 465  
DB 427 CAGGCGCAGATGTGTCACAGGCGCATGACCCCGCACCTGAACGCTGGTGAAGTG 486  
QY 466 ATCGAGGAGAAAGGCTTACGCGCGGAGGTATCCCATGTTACCGGCTGAGCGAGGCG 525  
DB 487 GTGAGAGAGAAAGGCTTACGCGCGGAGGTATCCCATGTTACGCGGCTGAGCGAGGCG 546  
QY 526 GCCACCGCGCGAGAGCTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGGCGCATG 585  
DB 547 GCCACCGCGCGAGAGCTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGGCGCATG 606  
QY 586 CAGATGCTGAAGAGACACCATCAACGAGAGGCGCGCGAGTGGGACCGGCTGACCCCGTG 645  
DB 607 CAGATGCTGAAGAGAGACCATCAACGAGAGGCGCGCGAGTGGGACCGGCTGACCCCGTG 666  
QY 646 CAGCGCGCGCGCGTGGCGCGCGCGAGATGCGCGGACCGCGCGGCGAGCATCGCGGCG 705  
DB 667 CAGCGCGCGCGCGATCGCGCGCGCGCGAGATGCGCGGACCGCGCGGCGAGCATCGCGGCG 726  
QY 706 GCCACGAGACCCCTGAGAGAGAGATGCGCTGGATGACGAAACCCCGCGTGGCGCGTG 765  
DB 727 ACCACGAGACCCCTGAGAGAGAGATGCGCTGGATGACGAAACCCCGCATCCCGCGTG 786  
QY 766 GCGGACATCTACAAGCGGTGATCATCTGGGCGTGAACAAGATCGTGGGATGTACAGC 825

DB 787 GCGGAGATCTACAAGCGGTGATCATCTGGGCGTGAACAAGATCGTGGGATGTACAGC 846  
QY 826 CCCGTGAGCATCTGAGCATCCGCGGAGGCCCAAGAGCCCTTCCGCGACTACGTGAGC 885  
DB 847 CCCACGAGCATCTGAGCATCCGCGGAGGCCCAAGAGCCCTTCCGCGACTACGTGAGC 906  
QY 886 CGCTTCTTAAGACCTGCGCGCGGAGCGGACCGGACCGGAGCGTGAAGAACTGATGACC 945  
DB 907 CGCTTCAACAAGACCTGCGCGCGTGAAGAGCGGACCGGAGCGTGAAGAACTGATGACC 966  
QY 946 GAGACCTGCTGTGAGAAAGCGCAACCGCGACTGCAAGACCATCTGCGGCTCTGCGC 1005  
DB 967 GAGACCTGCTGTGAGAAAGCGCAACCGCGACTGCAAGACCATCTGAGGCTCTGCGC 1026  
QY 1006 CCGCGCGCGCGCGTGAAGAGATGATGACCGCTGCGAGGCGGTGGGCGCGCGCGCAC 1065  
DB 1027 CCGCGCGCGCGCGTGAAGAGATGATGACCGCTGCGAGGCGGTGGGCGCGCGCGCAC 1086  
QY 1066 AAGGCGCGGTGCTGCGCGAGGCGGATGAGCCAGG---CCAACAGCTGAACATCATGATG 1122  
DB 1087 AAGGCGCGGTGCTGCGCGAGGCGGATGAGCCAGTGAACGAACCGGCGACATCATGATG 1146  
QY 1123 CAGAAGGCACTTCAAGGCGCGCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAG 1182  
DB 1147 CAGCGCGCGCACTTCCGCAACGAGCGGAAAGCCGTCAAGTCTTCAACTGCGGCAAGAG 1206  
QY 1183 GCGCACATCGCCAAAGACTGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGCAAG 1242  
DB 1207 GCGCACACCGCGCAAGACTGCGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGCGC 1266  
QY 1243 GAGGCGCACGATGAAGACTGCAACGAGCGCGGCAACTTCTTGGGCAAGATCTG 1302  
DB 1267 GAGGCGCACGATGAAGACTGCAACGAGCGCGGCAACTTCTTGGGCAAGATCTG 1326  
QY 1303 CCGACGCAAGAGGCGCGCGCGCAACTTCTGCAAGAACGCGAGCGAGCGCGCGCGCC 1362  
DB 1327 CCGAGCTAACAAGGCGCGCGCGCAACTTCTGCAAGAGCG-----C 1368  
QY 1363 ACCGTGCGCACCGCGCGCGCGCGGAGCTTCCGCTTGAAGAGACCAACCGCGCGCC 1422  
DB 1369 CCGAGCGCGCACCGCGCGCGCGCGGAGAGCTTCCGCTTGGCGAGAGAAAGCAACCGCC 1428  
QY 1423 AAGCAGAGCGCGCAAGAGCGCGGAGCGCTTACCGGAGCGCGCTGACCGCGCTGCGAGCTG 1482  
DB 1429 AGCCAGAAAGAGAGCGCGCATGCAACAAGAGCTGTACCGCTGACCGAGCTGCGAGCTG 1488  
QY 1483 TTGCGGAGCGCGCGCGCGCGTGAAGCGCACTTA 1509  
DB 1489 TTGCGGAGCGCGCGCGCGAGAGCGCACTTA 1515

Search completed: March 12, 2005, 21:18:23  
Job time : 6741.44 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 12, 2005, 16:41:47 ; Search time 266.651 Seconds  
(without alignments)  
9259.839 Million cell updates/sec

Title: US-09-475-704A-4  
Perfect score: 1509  
Sequence: 1 atgggccccgcgcgcacgcat.....gcggccccctgagccagtaa 1509

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1201	79.6	1515	4	US-09-475-515-4 Sequence 4, Appli
2	1201	79.6	4472	4	US-09-475-515-75 Sequence 75, Appli
3	1201	79.6	4608	4	US-09-475-515-76 Sequence 76, Appli
4	1201	79.6	4689	4	US-09-475-515-74 Sequence 74, Appli
5	1201	79.6	4766	4	US-09-475-515-73 Sequence 73, Appli
6	1199	79.5	2031	4	US-09-475-515-7 Sequence 7, Appli
7	1117.8	74.1	1853	4	US-09-475-515-5 Sequence 5, Appli
8	1117.8	74.1	1865	4	US-09-475-515-78 Sequence 78, Appli
9	1117.8	74.1	4319	4	US-09-475-515-79 Sequence 79, Appli
10	1117.8	74.1	1268	4	US-09-475-515-9 Sequence 9, Appli
11	1049.2	69.5	1521	4	US-09-952-060-27 Sequence 27, Appli
12	971.4	64.4	1532	4	US-09-974-702-1 Sequence 1, Appli
13	971.4	64.4	1532	4	US-09-974-702-1 Sequence 1, Appli
14	971.4	64.4	1532	4	US-09-974-702-1 Sequence 1, Appli
15	971.4	64.4	37474	4	US-09-952-060-25 Sequence 25, Appli
16	968.4	64.2	4053	4	US-09-952-060-34 Sequence 34, Appli
17	930.2	61.6	4307	4	US-09-936-572-2 Sequence 2, Appli
18	929.4	61.6	4307	4	US-09-936-572-2 Sequence 2, Appli
19	925.4	61.3	4327	4	US-09-936-572-14 Sequence 14, Appli
20	925.4	61.3	4353	4	US-09-936-572-13 Sequence 13, Appli
21	925.4	61.3	4642	4	US-09-936-572-12 Sequence 12, Appli
22	925.4	61.3	9772	4	US-09-552-950-5 Sequence 5, Appli
23	916.6	60.7	1482	4	US-09-818-443-4 Sequence 4, Appli
24	913.6	60.5	1479	4	US-09-952-060-32 Sequence 32, Appli
25	736.6	48.8	8366	4	US-09-872-733A-6 Sequence 6, Appli
26	734.2	48.7	1496	4	US-09-184-418C-82 Sequence 82, Appli
27	734.2	48.7	8972	4	US-09-184-418C-9 Sequence 9, Appli

28	730.4	48.4	1476	4	US-09-991-258-4 Sequence 4, Appli
29	730.4	48.4	12523	4	US-09-991-258-1 Sequence 1, Appli
30	723.8	48.0	4338	4	US-09-872-733A-1 Sequence 1, Appli
31	712.8	47.2	9913	4	US-09-827-688-11 Sequence 11, Appli
32	709.6	47.0	1485	4	US-09-184-418C-73 Sequence 73, Appli
33	709.6	47.0	9010	4	US-09-184-418C-8 Sequence 8, Appli
34	708	46.9	1476	4	US-09-184-418C-100 Sequence 100, App
35	708	46.9	8959	4	US-09-184-418C-11 Sequence 11, Appli
36	663.6	44.0	1486	4	US-09-184-418C-38 Sequence 38, Appli
37	663.6	44.0	8992	4	US-09-184-418C-4 Sequence 4, Appli
38	651.8	43.2	8968	4	US-09-184-418C-1 Sequence 1, Appli
39	649.4	43.0	1479	4	US-09-184-418C-12 Sequence 12, Appli
40	642.2	42.6	7399	2	US-08-418-848A-9 Sequence 9, Appli
41	642.2	42.6	9709	2	US-08-188-583-5 Sequence 5, Appli
42	642.2	42.6	9709	3	US-08-388-353-1 Sequence 1, Appli
43	642.2	42.6	9709	3	US-08-488-551B-1 Sequence 15, Appli
44	642.2	42.6	9709	3	US-09-309-572-15 Sequence 15, Appli
45	642.2	42.6	9709	4	US-09-718-096-15 Sequence 15, Appli

ALIGNMENTS

RESULT 1  
US-09-475-515-4  
; Sequence 4, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475, 515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1515  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-Gag  
US-09-475-515-4

Query Match 79.6%; Score 1201; DB 4; Length 1515;  
Best Local Similarity 88.5%; Pred. No. 2.8e-173;  
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;

QY	1	ATGGGCGCCCGCGCAGCATCTGCGCGGCGAGAGCTGAGCACTGGGAGAGATCCGC	60
DB	7	ATGGGCGCCCGCGCAGCGTGTGAGCGCGCGAGCTGAGCACTGGGAGAGATCCGC	66
QY	61	CTGGCGCCCGCGCGCAGAGCACTACATGCTGAAGCACCTGTGTGGCCAGCCGCGAG	120
DB	67	CTGGCGCCCGCGCGCAGAGCACTACATGCTGAAGCACCTGTGTGGCCAGCCGCGAG	126
QY	121	CTGGAGGCGCTTCGCGCTGAACCCCGCGCTGTGAGAGACCGCGAGGCGTGAAGCATC	180
DB	127	CTGGAGGCGCTTCGCGCTGAACCCCGCGCTGTGAGAGACCGCGAGGCGTGAAGCATC	186
QY	181	ATGAAGCAGCTGCAGCCCGCGCTGAGAGACCGAGAGAGCTGCAGAGCTGTACAAC	240
DB	187	CTGGGCGAGCTGCAGCCCGCGCTGAGAGACCGAGAGAGCTGCAGAGCTGTACAAC	246

```
QY 241 ACCGTGGCCACCCTGTACTGCGTGCAAGCCCGGCATCGAGGTCCGGCAACCAAGAGAGCC 300
Db 247 ACCGTGGCCACCCTGTACTGCGTGCAAGCCCGGCATCGAGGTCCGGCAACCAAGAGAGCC 306
QY 301 CTGACACAGATCGAGAGAGCAAGAACTCTCCAGCAGAAAGACCAGAGGCCAAGAG 360
Db 307 CTGACACAGATCGAGAGAGCAAGAACTCTCCAGCAGAAAGACCAGAGGCCAAGAG 366
QY 361 GCGGACGGCA-----AGGTGAGCCAGAACTACCCCATGTGACAGACTG 405
Db 367 GCGGACGGCAAGCCGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATGTGACAGACTG 426
QY 406 CAGGCGCAGATGCTGCAACAGGCCATCGACCCCGCAGCCCTGAAAGCTGGTGAAGTG 465
Db 427 CAGGCGCAGATGCTGCAACAGGCCATCGACCCCGCAGCCCTGAAAGCTGGTGAAGTG 486
QY 466 ATCGAGAGAAAGGCTTTCAGCCCGCAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGC 525
Db 487 GTGAGAGAGAAAGGCTTTCAGCCCGCAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGC 546
QY 526 GCCACCCCGCAGGACCTTGAACACAGATGTTGAACACCGTGGGGCGGCCACAGGCCCGCATG 585
Db 547 GCCACCCCGCAGGACCTTGAACACAGATGTTGAACACCGTGGGGCGGCCACAGGCCCGCATG 606
QY 586 CAGATGCTGAAGAGACACCATCAACGAGAGAGGCCCGCAGAGTGGGACCGCTGCACCCCGTG 645
Db 607 CAGATGCTGAAGAGAGACCATCAACGAGAGAGGCCCGCAGAGTGGGACCGCTGCACCCCGTG 666
QY 646 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705
Db 667 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
QY 706 GCCACGACACCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
Db 727 ACCACGACACCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
QY 766 GGCGACATCTACAAGCGGTGATCATCTCTGGGCTGAAACAAGATGCTGGGATGTAACAGC 825
Db 787 GGCGACATCTACAAGCGGTGATCATCTCTGGGCTGAAACAAGATGCTGGGATGTAACAGC 846
QY 826 CCGGTGAGCATCCTGATCATCCGCGCAGAGGCCCTTCCGCGACTAGCTGAGC 885
Db 847 CCGGTGAGCATCCTGATCATCCGCGCAGAGGCCCTTCCGCGACTAGCTGAGC 906
QY 886 CGCTTCTTCAAGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945
Db 907 CGCTTCTTCAAGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966
QY 946 GAGACCTTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005
Db 967 GAGACCTTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
QY 1006 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1065
Db 1027 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1086
QY 1066 AAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
Db 1087 AAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1146
QY 1123 CAGAAAGAGCACTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182
Db 1147 CAGCGCGCGCACTTCCGCAACGAGCGGAGAGACCGTCAAGTGTCTTCACTGCGCGCAAGAG 1206
QY 1183 GCGCACATCGCCAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242
Db 1207 GCGCACACCGCGAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1266
QY 1243 GAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302
Db 1267 GAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1326
QY 1303 CCCAGCCACAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1362
```

```
Db 1327 CCCAGCTACAAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368
QY 1363 ACCGTGGCCACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1422
Db 1369 CCCAGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1428
QY 1423 AAGCAGAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482
Db 1429 AGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
QY 1483 TTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1509
Db 1489 TTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1515

RESULT 2
US-09-475-515-75
; Sequence 75, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGED, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 4472
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; US-09-475-515-75

Query Match          79.6%; Score 1201; DB 4; Length 4472;
Best Local Similarity 88.5%; Pred. No. 2.8e-173;
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;

QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db 2939 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2998
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 2999 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3058
QY 121 CTGAGGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 3059 CTGAGGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3118
QY 181 ATGAAGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 3119 CTGAGGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3178
QY 241 ACCGTGGCCACCCTGTAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 3179 ACCGTGGCCACCCTGTAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3238
QY 301 CTGACACAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 3239 CTGACACAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3298
```

```
QY 361 GCCGACGCA-----AGGTGAGCCAGAACTACCCCATCGTGCAGAACCTG 405
Db 3299 GCCGCCGCGCACCGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTG 3358
QY 406 CAGGCGCAGATGTGCAACAGGCCATGACCCCGCACCCCTGAACGCTGGTGAAGTG 465
Db 3359 CAGGCGCAGATGTGCAACAGGCCATGACCCCGCACCCCTGAACGCTGGTGAAGTG 3418
QY 466 ATGAGAGAAAGGCTTCAGCCCGCAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCG 525
Db 3419 GTGAGAGAGAAAGGCTTCAGCCCGCAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCG 3478
QY 526 GCCACCCCCCAGGACCTGGAACAGATGTTGAACACCCGTGGCGGCCACAGCGCCCATG 585
Db 3479 GCCACCCCCCAGGACCTGGAACAGATGTTGAACACCCGTGGCGGCCACAGCGCCCATG 3538
QY 586 CAGATGCTGAAGAGCAACCATCAACGAGAGGCGCCGAGTGGGACCCGCTGCACCCCGTG 645
Db 3539 CAGATGCTGAAGAGCAACCATCAACGAGAGGCGCCGAGTGGGACCCGCTGCACCCCGTG 3598
QY 646 CAGGCGCGCGCGCGTGGCGCGCGCGCGCAAGTGGCGGACCCCGCGGAGACATCGCGCGC 705
Db 3599 CAGGCGCGCGCGCGTGGCGCGCGCGCGCGCAAGTGGCGGACCCCGCGGAGACATCGCGCGC 3658
QY 706 GCCACCAAGACCCCTGCAAGAGAGAGATGCGCTGATGACCAAGCAACCCCGCGTGGC 765
Db 3659 ACCACCAAGACCCCTGCAAGAGAGAGATGCGCTGATGACCAAGCAACCCCGCGTGGC 3718
QY 766 GGGCAGATCTACAAGCGGTGGATCATCTGGGGCTGAACAAGATCGTGGGATGTACAGC 825
Db 3719 GGGCAGATCTACAAGCGGTGGATCATCTGGGGCTGAACAAGATCGTGGGATGTACAGC 3778
QY 826 CCCGTGAGCATCTGGAACATCCCGCGAGGCGCCCAAGAGCCCTTCCCGGACTACGTGAGC 885
Db 3779 CCCGTGAGCATCTGGAACATCCCGCGAGGCGCCCAAGAGCCCTTCCCGGACTACGTGAGC 3838
QY 886 CGCTTCTTCAAGACCCCTGCGCGCGCGCGAGGAGCCCAAGAGCGTGAAGAACTGATGAC 945
Db 3839 CGCTTCTTCAAGACCCCTGCGCGCGCGCGAGGAGCCCAAGAGCGTGAAGAACTGATGAC 3898
QY 946 GAGACCCCTGCTGTGTGAGAGAACGCCAACCCCGAGCTGCAAGAACCATCTGCGCGCTG 1005
Db 3899 GAGACCCCTGCTGTGTGAGAGAACGCCAACCCCGAGCTGCAAGAACCATCTGCGCGCTG 3958
QY 1006 CCGCGCGCGCGCGCTGAGAGAGATGATGACCGCTGCGAGGCGCGTGGCGCGCGCGCGC 1065
Db 3959 CCGCGCGCGCGCGCTGAGAGAGATGATGACCGCTGCGAGGCGCGTGGCGCGCGCGCGC 4018
QY 1066 AAGGCGCGCGCTGTGGCGCGAGGCGGATGAGCCAGG--CCAAGCGGTGAACATCATGATG 1122
Db 4019 AAGGCGCGCGCTGTGGCGCGAGGCGGATGAGCCAGGCGGAGCCCGCGGAGCCATCATGATG 4078
QY 1123 CAGAAAGAGCAACTTCAAGGGCGCGCGCGCAACGTCAAGTGTCTTCACTGCGGCAAGAG 1182
Db 4079 CAGCGCGCGCAACTTCCGCAACAGCGGAAAGACCGTCAAGTGTCTTCACTGCGGCAAGAG 4138
QY 1183 GGCCACATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGCAGGAG 1242
Db 4139 GGCCACATGCGCAAGAACTGCGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGCAGGAG 4198
QY 1243 GAGGCGCACAGATGAGAGACTGACCGAGCGCGCAAGCTTCTGCGGCAAGATCTGG 1302
Db 4199 GAGGCGCACAGATGAGAGACTGACCGAGCGCGCAAGCTTCTGCGGCAAGATCTGG 4258
QY 1303 CCCAGCCACAAGGGCGCGCGCGCGCAACTTCTGCGAGAACCGGAGCGAGCCCGCGCGCC 1362
Db 4259 CCCAGCCACAAGGGCGCGCGCGCGCAACTTCTGCGAGAACCGGAGCGAGCCCGCGCGCC 4300
QY 1363 ACCGTGCGCACCGCGCGCGCGCGCGAGAGCTTCCGCTTGAAGAGAACCCCGCGCGCC 1422
Db 4301 CCCGAGCGCACCGCGCGCGCGCGAGAGAGCTTCCGCTTGAAGAGAACCCCGCGCGCC 4360
```

```
QY 1423 AAGCAGAGCCCAAGAGACCGCGAGCCCTACCGCGAGCCCTGACCGCCCTGCGCAGCTG 1482
Db 4361 AGCCAGAAAGCAGAGAGCCCATGCAAGAGAGCTGTATCCCTGACCAAGCTTGCAGCCTG 4420
QY 1483 TTGGCAGCGCGCGCGCGCTGAGCCAGTAA 1509
Db 4421 TTGGCAGCGAGCCCGCAGCAGCCAGTAA 4447
```

## RESULT 3

```
US-09-475-515-76
; Sequence 76, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDS, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 76
; LENGTH: 4608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp160.modsF162.delV2.gag.modsF2
US-09-475-515-76
```

```
Query Match 79.6%; Score 1201; DB 4; Length 4608;
Best Local Similarity 88.5%; Pred. No. 2.8e-173;
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;

QY 1 ATGGGGCGCGCGCGCGCGCATCTGCGCGCGCGGAGAACTGGACAAAGTGGGAGAAATCCGC 60
Db 3075 ATGGGGCGCGCGCGCGCGCATCTGCGCGCGCGGAGAACTGGGAGAAAGTGGGAGAAATCCGC 3134
QY 61 CTGGCGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCCAGCGCGGAG 120
Db 3135 CTGGCGCGCGCGCGCGCAAGAGAGTACAGCTGAAGCACTGTGTGGGCCAGCGCGGAG 3194
QY 121 CTGAGGGGCTTCCGCTGAACCCCGCGCTGCTGAGAACCGCGGAGGGCTGCAAGCATG 180
Db 3195 CTGAGGGGCTTCCGCTGAACCCCGCGCTGCTGAGAACCGCGGAGGGCTGCGCGGAGATC 3254
QY 3195 CTGAGGGGCTTCCGCTGAACCCCGCGCTGCTGAGAACCGCGGAGGGCTGCGCGGAGATC 3254
QY 181 ATGAAGCAGCTGAGCGCGCGCTGCAAGACCGGCAACGAGAGCTGCGGAGCTGTACAAC 240
Db 3255 CTGGGCCAGCTGAGCGCGCGCTGCAAGACCGGCAACGAGAGCTGCGGAGCTGTACAAC 3314
QY 241 ACCGTGGCCACCTGTACTGCTGCAACCGCGCATGAGTCCGCGGACCAAGAGGGCC 300
Db 3315 ACCGTGGCCACCTGTACTGCTGCAACCGCGCATGAGTCCGCGGACCAAGAGGGCC 3374
QY 301 CTGGAACAAGATGAGAGAGAGAGCAACAAGTCCAGCAGAAAGACCCAGAGGCCAAGAG 360
Db 3375 CTGGAACAAGATGAGAGAGAGAGCAACAAGTCCAGCAGAAAGACCCAGAGGCCAAGAG 3434
QY 361 GCCGACGCA-----AGGTGAGCGAAGAACTACCCCATCGTGCAAGAACTG 405
Db 3435 GCCGCGCGCACCGCGCAACAGACGCCAGGTGAGCGAAGAACTACCCCATCGTGCAAGAACTG 3494
QY 406 CAGGCGCAGATGTGCAACAGGCCATGACCCCGGCAACCTGAGCGCTGGGTGAAGGTG 465
```

Db	3495	CAGGGCCAGATGGTGCAACGAGCCATCAGCCCCCGCACTTGAACGCTGGTGAAGTG	3554
QY	466	ATCGAGGAGAAAGGCGCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGC	525
Db	3555	GTGAGGAGAAAGGCGCTTCAGCCCCGAGGTGATCCCATGTTCAAGCGCGCTGAGCGAGGGC	3614
QY	526	GCCACCCCCCAGGACCTTGAACAAGATGTTGAACACCGTGGCGGCGCACAGGCGCCATG	585
Db	3615	GCCACCCCCCAGGACCTTGAACAAGATGTTGAACACCGTGGCGGCGCACAGGCGCCATG	3674
QY	586	CAGATGCTGAAGSACACCATCAACGAGAGAGGCGCGGAGTGGAGCTGCTGCACCCCGTG	645
Db	3675	CAGATGCTGAAGSACACCATCAACGAGAGAGGCGCGGAGTGGAGCTGCTGCACCCCGTG	3734
QY	646	CAGCGCGGCGCGCGTGGCGCGCGCGCGCGAGATGCGCGACCCCCCGCGGACCGACATCGCGGC	705
Db	3735	CAGCGCGGCGCGCGTGGCGCGCGCGCGCGAGATGCGCGACCCCCCGCGGACCGACATCGCGGC	3794
QY	706	GCCACGAGCACCCTGCGAGGAGAGATCGCCTGATGACCAAGATCGTGGCATGTACAGC	765
Db	3795	ACCACGAGCACCCTGCGAGGAGAGATCGCCTGATGACCAAGATCGTGGCATGTACAGC	3854
QY	766	GGCGCATCTCAAGACCGGTGATCATCTGGCGCTGAACAAGATCGTGGCATGTACAGC	825
Db	3855	GGCGCATCTCAAGACCGGTGATCATCTGGCGCTGAACAAGATCGTGGCATGTACAGC	3914
QY	826	CCCGTGAGCATCTGAGACATCCGCGAGAGGCGCCCAAGAGAGCCTTCCGCACTACGTGAC	885
Db	3915	CCCAACGACATCTGAGACATCCGCGAGAGGCGCCCAAGAGAGCCTTCCGCACTACGTGAC	3974
QY	886	CGCTTCTTCAAGACCCCTGCGCGCGAGAGCGCACCCAGAGCTGAAAGACTGATGACC	945
Db	3975	CGCTTCTTCAAGACCCCTGCGCGCTGAGAGCGCACCCAGAGCTGAAAGACTGATGACC	4034
QY	946	GAGACCTCTGCTGGTGAGAACCGCCAAACCCCGACTGCAAGACATCTGCGCGCTCTCGC	1005
Db	4035	GAGACCTCTGCTGGTGAGAACCGCCAAACCCCGACTGCAAGACATCTGCGCGCTCTCGC	4094
QY	1006	CCCGGCGGCACCTCTGGAGAGATGATGACCGCTTCCAGGCGGTGGCGGCGCGCCAC	1065
Db	4095	CCCGGCGGCACCTCTGGAGAGATGATGACCGCTTCCAGGCGGTGGCGGCGCGCCAC	4154
QY	1066	AAGGCGCGGCTGCTGGCGCGAGGCGATGAGCCAGG--CCAAACGCTGAACATCATGATG	1122
Db	4155	AAGGCGCGGCTGCTGGCGCGAGGCGATGAGCCAGGCTGACCAACCCGCGCACATCATGATG	4214
QY	1123	CAGAAAGCAACTTCAAGGCGCGCGCGCGCAACGTCAAGTCTTCACTGCGCAAGAG	1182
Db	4215	CAGCGCGCAACTTCCGCAACGAGCGGAAGAACCGTCAAGTCTTCACTGCGCAAGAG	4274
QY	1183	GCGCAATCGCCCAAGAACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAG	1242
Db	4275	GCGCAATCGCCCAAGAACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAG	4334
QY	1243	GAGGCGCACCATGAAGGACTGCAACGAGGCGCGAGGCAACTTCTGGGCAAGATCTGG	1302
Db	4335	GAGGCGCACCATGAAGGACTGCAACGAGGCGCGAGGCAACTTCTGGGCAAGATCTGG	4394
QY	1303	CCCAGCCACAAGGCGCGCGCGCAACTTCTTGAGAACCGCAGCGCGCGCGCGCGCG	1362
Db	4395	CCCAGCTACAAGGCGCGCGCGCGCAACTTCTTGAGAGCGG-----C	4436
QY	1363	ACCGTGCCCAACGCGCGCGCGCGCGAGAGCTTCCGTTGAGAGAGACACCCCGCGCGCG	1422
Db	4437	CCCGAGCCCAACGCGCGCGCGCGCGAGAGAGCTTCCGTTGCGAGAGAGAACACACCGCG	4496
QY	1423	AAGCAGGAGCCCAAGGACCGCGAGCCCTACCGCGAGCGCGCTGACCGCGCTGCGAGCTG	1482
Db	4497	AAGCAGGAGCCCAAGGACCGCGAGCCCTACCGCGAGCGCGCTGACCGCGCTGCGAGCTG	4556
QY	1483	TTTGGGAGCGGCGCGCTGAGCCAGTAA	1509
Db	4557	TTTGGGAGCGGCGCGCTGAGCCAGTAA	4583

```

RESULT 4
US-09-475-515-74
; Sequence 74, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 74
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp160.modsf162.gag.modsf2
US-09-475-515-74

```

Query Match	79.6%	Score 1201	DB 4	Length 4689
Best Local Similarity	88.5%	Pred. No. 2.8e-173		
Matches 1351	Conservative 0	Mismatches 140	Indels 36	Gaps 3
QY 1	ATGGCGCCCGCCGACGATCTCGCGCGGAGAGAGCTGGACAAAGTGGAGAGATCCGC			60
Db 3156	ATGGCGCCCGCCGACGCTGTGAGCGGCGGAGCTGGACAAAGTGGAGAGATCCGC			3215
QY 61	CTGCGCCCGCCGCGCAAGAACACTACATGCTGAAGCACCTGTGTGGGCGACCGCGAG			120
Db 3216	CTGCGCCCGCCGCGCAAGAGAGTACAGCTGAAGCACATCGTGTGGGCGACCGCGAG			3275
QY 121	CTGAGAGGCTTCGCGCTGAACCCCGGCTGTGAGACCGCCGAGGGCTGCAAGCAGATC			180
Db 3276	CTGAGAGGCTTCGCGCTGAACCCCGGCTGTGAGACCGAGGGCTGCGCAGATC			3335
QY 181	ATGAAGCAGCTGCAGCGCCGCTGTGCAGACCGGACCGAGAGCTGCGCAGCTGTACAAC			240
Db 3336	CTGGGCGCAGCTGCAGCGCCAGCTGTGCAGACCGGACCGAGAGCTGCGCAGCTGTACAAC			3395
QY 241	ACCGTGGCCACCCCTGTACTGCGGTGCAGCGCCGCGCATCGAGTCCGCGACCAAGAGGCC			300
Db 3396	ACCGTGGCCACCCCTGTACTGCGGTGCAGCGCGCATCGAGTCCAGGACCAAGAGAGGCC			3455
QY 301	CTGGAACAAGATGAGAGAGAGCAAGCAAGTCCAGAGAAAGACCAGAGGCCAAGAGAG			360
Db 3456	CTGGAAGAAGATGAGAGAGAGCAAGCAAGTCCAGAGAAAGGCCAGAGGCCGCCGCC			3515
QY 361	GCCGACGGCA-----AGGTAGCCAGAACTACCCCATCTGTGCAGAACTG			405
Db 3516	GCCGCGGGCACCGGCAACAGCAGCGCAGGTGAGCCAGAACTACCCCATCTGTGCAGAACTG			3575
QY 406	CAGGGCAGATGTGTGCAACAGGCCATCAGCCCCCGCACCTGAAACGCTGGTGAAGGTG			465
Db 3576	CAGGGCAGATGTGTGCAACAGGCCATCAGCCCCCGCACCTGAAACGCTGGTGAAGGTG			3635
QY 466	ATCGAGAGAGAGGCTTCAGCCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGC			525
Db 3636	GTGAGAGAGAGGCTTCAGCCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGC			3695
QY 526	GCCACCCCCCAGACCTGAACACAGATGTTGAACACCGTGGGCGGCCACAGGCCGCCATG			585

Db 3696 GCCACCCCGCAGACCTGAAACAGATGTTGAACACCGTGGCGGCCACACGAGCCGCCATG 3755  
QY 586 CAGATGCTGAAGAGACACCATCAACGAGAGGCGCCGAGTGGGACCGCTGCACCCCGTG 645  
Db 3756 CAGATGCTGAAGAGACCATCAACGAGAGGCGCCGAGTGGGACCGCTGCACCCCGTG 3815  
QY 646 CAGGCGGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705  
Db 3816 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3875  
QY 706 GCCACGACACCTGACAGAGCAGATCGCTGGATGACCAACACCCCGCTGCGCGTG 765  
Db 3876 ACCACGACACCTGACAGAGCAGATCGCTGGATGACCAACACCCCGCTGCGCGTG 3935  
QY 766 GCGGACATCTACAAGCGGTGATCATCTGCGGCTGAAACAAGATCGTGGGATGTACAGC 825  
Db 3936 GCGGAGATCTACAAGCGGTGATCATCTGCGGCTGAAACAAGATCGTGGGATGTACAGC 3995  
QY 826 CCGGTGACATCTGACATCCGCGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGAC 885  
Db 3996 CCCACGACATCTGACATCCGCGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGAC 4055  
QY 886 CGCTTCTTCAAGACCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945  
Db 4056 CGCTTCTTCAAGACCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4115  
QY 946 GAGACCTCTGCTGTGACAGACGCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGC 1005  
Db 4116 GAGACCTCTGCTGTGACAGACGCAACCCCGACTGCAAGACCATCTGAGGCTCTCGGC 4175  
QY 1006 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1065  
Db 4176 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4235  
QY 1066 AAGGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122  
Db 4236 AAGGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4295  
QY 1123 CAGAAAGCAACTTCAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182  
Db 4296 CAGCGCGCGCAACTTCCGCAACCGCGGAGAACCGTCAAGTGTCTCAACTGCGCGCAAGAG 4355  
QY 1183 GGGCACAATCGCCAAAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242  
Db 4356 GGGCACAATCGCCAAAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4415  
QY 1243 GAGGGCCACAGATGAAGACTGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302  
Db 4416 GAGGGCCACAGATGAAGACTGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4475  
QY 1303 CCGAGCCACAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1362  
Db 4476 CCGAGCTACAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4517  
QY 1363 ACCGTGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1422  
Db 4518 CCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4577  
QY 1423 AAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1482  
Db 4578 AAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4637  
QY 1483 TTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1509  
Db 4638 TTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4664

RESULT 5  
US-09-475-515-73  
; Sequence 73, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:

APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Ying  
APPLICANT: HARTOG, Karin  
APPLICANT: LIU, Hong  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 73  
LENGTH: 4766  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-09-475-515-73

Query Match 79.6%; Score 1201; DB 4; Length 4766;  
Best Local Similarity 88.5%; Pred. No. 2.8e-173;  
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;  
QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 3233 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3292  
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 3293 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3352  
QY 121 CTGAGGGCTTGGCGCTGAACCCCGCGCTGCTGAGACCGCGGAGGCTGCAAGCATC 180  
Db 3353 CTGAGGGCTTGGCGCTGAACCCCGCGCTGCTGAGACCGCGGAGGCTGCGCGCATC 3412  
QY 181 ATGAAGCAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 3413 CTGGGCCAGCTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3472  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 3473 ACCGTGGCCACCTGTACTGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3532  
QY 301 CTGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 3533 CTGGAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3592  
QY 361 GCCAGCGCA-----AGTGAGCCAGAACTACCCCATCTGTGACAGAACTG 405  
Db 3593 GCCCGCGCACCGCGCAACAGCAGCGAGTGAGCGCAAGTAACCCCATCTGTGACAGAACTG 3652  
QY 406 CAGGCGCAGATGTGACACGAGCGCATCAGCCCGCGCACCCCTGAAGCGCTGGTGAAGTG 465  
Db 3653 CAGGCGCAGATGTGACACGAGCGCATCAGCCCGCGCACCCCTGAAGCGCTGGTGAAGTG 3712  
QY 466 ATGAGAGAGAGGCTTCAAGCGCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCG 525  
Db 3713 GTGAGAGAGAGGCTTCAAGCGCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCG 3772  
QY 526 GCCACCCCGCAGGACTGAACAGATGTTGAACACCGTGGCGCGCACAGGCGCGCATG 585  
Db 3773 GCCACCCCGCAGGACTGAACAGATGTTGAACACCGTGGCGCGCACAGGCGCGCATG 3832  
QY 586 CAGATGCTGAAGGACACCATCAACGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645  
Db 3833 CAGATGCTGAAGGAGACCATCAACGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3892

```
QY 646 CAGGCCGGCCCGTGGCCCCCGGGCCAGATGCGGCAACCCCGGGCAGCGACATGCGCGGC 705
Db 3893 CAGCGCGGCCCATCGCCCCCGGCCAGATGCGGAGAGCCCGCGCAGCGACATGCGCGGC 3952
QY 706 GCCACCAAGCACTTGCAGAGGAGAGATGCGCTGGATGACCAACACCCCGCTGCGCGCTG 765
Db 3953 ACCACCAAGCACTTGCAGAGGAGAGATGCGCTGGATGACCAACACCCCGCTGCGCGCTG 4012
QY 766 GCGGACATCTACAAGCGGTGATCATCTTGGGCTTGAACAAGATGCTGGGATGTACAGC 825
Db 4013 GCGGAGATCTACAAGCGGTGATCATCTTGGGCTTGAACAAGATGCTGGGATGTACAGC 4072
QY 826 CCGGTGAGCATCTTGAACATCCGCCAGAGGCCCCCAAGAGCCCTTCCGCGACTACGTGAC 885
Db 4073 CCGACCAAGCATCTTGAACATCCGCCAGAGGCCCCCAAGAGCCCTTCCGCGACTACGTGAC 4132
QY 886 CGCTTCTTAAAGACCTTGCGCGCGCCGAGAGGCGCACCCAGGACGTGAAGAACTGATGACC 945
Db 4133 CGCTTCTTAAAGACCTTGCGCGCGCTGAGAGGCGCACCCAGGACGTGAAGAACTGATGACC 4192
QY 946 GAGACCTCTGTGTGTCAGAAACGCCAACCCCGACTGCAAGAACCATCTGCGCGCTCTCGGC 1005
Db 4193 GAGACCTCTGTGTGTCAGAAACGCCAACCCCGACTGCAAGAACCATCTGAAAGGCTCTCGGC 4252
QY 1006 CCGGCGCGCAACCTTGAGAGATGATGACCGCTGCGAGGGCGTGGGCGCGCCCGCGCAC 1065
Db 4253 CCGGCGCGCAACCTTGAGAGATGATGACCGCTGCGAGGGCGTGGGCGCGCCCGCGCAC 4312
QY 1066 AAGCGCGCGCTGTGCGCGCAGCGATGAGCCAGG--CCAACAGCGTGAACATCATGATG 1122
Db 4313 AAGCGCGCGCTGTGCGCGCAGCGATGAGCCAGGTGACGAAACCCGCGCAACATCATGATG 4372
QY 1123 CAGAAGCACTTCAAGGGGCCCCCGGGCGCAACGTCAAGTGTCTTCAACTGCGGCAAGAG 1182
Db 4373 CAGCGCGCGCACTTCCGCAACCAAGCGGAAAGACCGTCAAGTGTCTTCAACTGCGGCAAGAG 4432
QY 1183 GGCCCACTGCGCAAGAACTGCGCGCGCCCCCGCAAGAAAGGGCTGCTGGAAGTGGCGCAAG 1242
Db 4433 GGCCCACTGCGCAAGAACTGCGCGCGCCCCCGCAAGAAAGGGCTGCTGGAAGTGGCGCGC 4492
QY 1243 GAGGCGCCACAGATGAAGACTGCAACCGCAGCGCCAGCCCACTTCTTGGGCAAGATCTGG 1302
Db 4493 GAGGCGCCACAGATGAAGACTGCAACCGCAGCGCCAGCCCACTTCTTGGGCAAGATCTGG 4552
QY 1303 CCCAGCCACAAGGCGCGCGCGCGCAACTTCTGTCAGAAACCGCAGCGACCGCGCGCGCC 1362
Db 4553 CCCAGCTTCAAGGCGCGCGCGCGCAACTTCTGTCAGAGCGG-----C 4594
QY 1363 ACCGTGCCCCACCGCGCGCGCGCGCGAGAGCTTCCGCTTTCAGAGAGACCAACCCCGCGCC 1422
Db 4595 CCGGAGCCCAACCGCGCGCGCGCGCGAGAGAGCTTCCGCTTTCGCGAGAGAAAGACCAACCC 4654
QY 1423 AAGCAGAGAGCCCAAGAGACCGCGAGCGCTACCGGAGCGCGCTGACCGCGCTGCGGAGCCTG 1482
Db 4655 AGCCAGAGAGAGAGAGAGCCCATGCAAGAGAGCTGTACCCCTGACCAAGCCTGCGGAGCCTG 4714
QY 1483 TTGCGCAGCGCGCGCGCTGAGCCAGTAA 1509
Db 4715 TTCGCAACGACCCCAAGCAGCCAGTAA 4741
```

## RESULT 6

```
US-09-475-515-7
; Sequence 7, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
```

```
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475, 515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-475-515-7
```

```
Query Match 79.5%; Score 1199; DB 4; Length 2031;
Best Local Similarity 88.5%; Pred. No. 5.6e-173;
Matches 1349; Conservative 0; Mismatches 140; Indels 36; Gaps 3;
```

```
QY 1 ATGGGCGCGCGCGCCAGCATCTGCGCGCGAGAAAGCTGGAACAAGTGGGAAAGATCCGC 60
Db 7 ATGGGCGCGCGCGCCAGCATCTGCGCGCGAGAGCTGGAACAAGTGGGAAAGATCCGC 66
QY 61 CTGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCGCGAG 120
Db 67 CTGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCGCGAG 126
QY 121 CTGAGGCGCTTTCGCGCTGAAACCCCGCGCTGTGAGACCGCGAGGCTGCAAGCAGATC 180
Db 127 CTGAGGCGCTTTCGCGCTGAAACCCCGCGCTGTGAGACCGAGGCTGCGCAGATC 186
QY 181 ATGAAGCAGCTGACGCGCGCGCTGACAGACCGGACCGAGAGCTGCGGAGCTGTACAA 240
Db 187 CTGGGCGAGCTGACGCGCGCGCTGACAGACCGGACCGAGAGAGCTGCGGAGCTGTACAA 246
QY 241 ACCGTGGCAACCTGTACTGCTGTCAGCGCGCGCATGAGGTCCGCAACCAAGAGGCG 300
Db 247 ACCGTGGCAACCTGTACTGCTGTCAGCGCGCGCATGAGGTCCGCAACCAAGAGGCG 306
QY 301 CTGGAACAAGTGAAGAGAGAGAGAGCAAGTCCAGCAGAAAGACCGAGGCGCAAGAG 360
Db 307 CTGGAACAAGTGAAGAGAGAGAGAGCAAGTCCAGAAAGAGAGCGCGCGCGCG 366
QY 361 GCCGACGGCA-----AGGTGAAGCAGAACTAACCCCATGTGCAAGAACTG 405
Db 367 GCCGCGCGCAACCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
QY 406 CAGGCGCAGATGTGTCAGAGGAGCATGACCCCGCGACCTTGAACGCTGGGTGAAGTG 465
Db 427 CAGGCGCAGATGTGTCAGAGGAGCATGACCCCGCGACCTTGAACGCTGGGTGAAGTG 486
QY 466 ATGAGAGAGAGGCGCTTCAAGCGCGCGAGGTATCCCAATGTTCAACCGCGCTGAGCGGG 525
Db 487 GTGAGAGAGAGGCGCTTCAAGCGCGCGAGGTATCCCAATGTTCAACCGCGCTGAGCGGG 546
QY 526 GCCACCGCGCGAGACCTGAACACAGATGTTGAACAACCGTGGGCGCGCAACAGCGCGCAT 585
Db 547 GCCACCGCGCGAGACCTGAACACAGATGTTGAACAACCGTGGGCGCGCAACAGCGCGCAT 606
QY 586 CAGATGCTGAAGAGACACATCAACGAGAGAGCGCGCGAGTGGAGACCGCTGACACCGGTG 645
Db 607 CAGATGCTGAAGAGAGACACATCAACGAGAGAGCGCGCGAGTGGAGACCGCTGACACCGGTG 666
QY 646 CAGGCGCGCGCGCGCGCGCGCGCGAGATGCGGAGACCGCGCGCGAGCATGCGCGGC 705
Db 667 CAGGCGCGCGCGCGCGCGCGCGCGAGATGCGGAGACCGCGCGCGAGCATGCGCGGC 726
QY 706 GCCACCAAGCACTTGCAGAGAGAGAGATGCGCTGATGACCAAGAAACCCCGTGGCGGTG 765
Db 727 ACCACCAAGCACTTGCAGAGAGAGAGATGCGCTGATGACCAAGAAACCCCGCATCCCGTG 786
```

QY	766	GGCGCATCTTCAAGCGGTGGATCATCTTGGGCGCTTGAA	CAAGATGCTGCGGATGTACAGC	825
Db	787	GGCGAGATCTTCAAGCGGTGGATCATCTTGGGCGCTTGAA	CAAGATGCTGCGGATGTACAGC	846
QY	826	CCCGTGAAGCATCCTTGGACATCCGCGCAGAGGGCCCC	CAAGAGACCCCTTCGCGACTACGTGGAC	885
Db	847	CCCAACGACATCCTTGGACATCCGCGCAGAGGGCCCC	CAAGAGACCCCTTCGCGACTACGTGGAC	906
QY	886	CGCTTCTTCAAGAACCTTCGCGCGCGCGACAGGGCC	ACCCAGGACGTGAAGAACTGGATGACC	945
Db	907	CGCTTCTTCAAGAACCTTCGCGCGCGCGACAGGGCC	ACCCAGGACGTGAAGAACTGGATGACC	966
QY	946	GAGACCTCTGTGTGTGACAGAACGCCAACCCCGA	CTGCAAGAACCATCTTGCAGCTCTCGGC	1005
Db	967	GAGACCTCTGTGTGTGACAGAACGCCAACCCCGA	CTGCAAGAACCATCTTGAAGGCTCTCGGC	1026
QY	1006	CCCGGCGCCACCCCTGAGAGAGATGATACCGCTTC	CAAGGCGCTGGCGGCCGCCAC	1065
Db	1027	CCCGGCGCCACCCCTGAGAGAGATGATACCGCTTC	CAAGGCGCTGGCGGCCGCCAC	1086
QY	1066	AAGGCGCGCGTGTGTGCGCGAGGCGATGAGCCAGG	---CCAACAGGTGAACATCATGATG	1122
Db	1087	AAGGCGCGCGTGTGTGCGCGAGGCGATGAGCCAGG	---CCAACAGGTGAACATCATGATG	1146
QY	1123	CAGAAAGAGCAACTTCAAGGGCCCCCGCGCAACGT	CAAGTGTCTTCACTGCGGCAAGAG	1182
Db	1147	CAGCGCGGCAACTTCCGCAACAGGGGGAAGACCGT	CAAGTGTCTTCACTGCGGCAAGAG	1206
QY	1183	GGCCACATCGCCCAAGAACTGCGCGCCCCCGCA	AGAAAGGCTGTGGAAGTGGCGCAAG	1242
Db	1207	GGCCACATCGCCCAAGAACTGCGCGCCCCCGCA	AGAAAGGCTGTGCGTGGCGCGCGC	1266
QY	1243	GAGGCGCACCAAGTGAAGAGACTGCAACGAGCGCA	GGCCAACTTCTTGGCAAGATCTGG	1302
Db	1267	GAGGCGCACCAAGTGAAGAGACTGCAACGAGCGCA	GGCCAACTTCTTGGCAAGATCTGG	1326
QY	1303	CCCAAGCCCAAGGGCGCGCCCGGCAACTTCTCTG	CAGAAACCGCAAGCGAGCCGCGCGCCCC	1362
Db	1327	CCCAAGCTACAAAGGCGCGCCCGGCAACTTCTCTG	CAGAGCGG-----C	1368
QY	1363	ACCGTCCCAACGCGCGCGCGCGCGAGAGCTTCCG	CTTGAAGAGACCAACCCCGCGCCCC	1422
Db	1369	CCCGAGCCCAACGCGCGCGCGCGCGAGAGCTTCCG	CTTGAAGAGAGAACCAACCCCGCGCCCC	1428
QY	1423	AAGCAGAGAGCCCAAGGACCGCGAGCCCTAACCG	GAGCCCTGACCGCCCTGCGCAGCCTG	1482
Db	1429	AAGCAGAGAGCCCAAGGACCGCATGCAAAAGAGCT	GTACCCCTGACCAAGCCTGCGCAGCCTG	1488
QY	1483	TTTCGGCAGCGGCGCCCTGAGCCAGT	1507	
Db	1489	TTTCGGCAACGACCCCAAGCAGCCAGT	1513	

```

RESULT 7
US-09-475-515-5
; Sequence 5, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30

```

```

; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1853
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: HIV-Gag-protease
US-09-475-515-5

```

Query Match	74.1%;	Score 1117.8;	DB 4;	Length 1853;
Best Local Similarity	85.1%;	Pred. No. 1e-160;		
Matches 1299;	Conservative	0;	Mismatches 192;	Indels 36;
				Gaps 3;

QY	1	ATGGGCGCCCGCCCAAGCATCTGCGCGGCGGAGAACTGGACAAGTGGAGAAATCCG	60
Db	7	ATGGGCGCCCGCCCAAGCGTGTGAGCGGCGGCGAGCTGGACAAGTGGAGAAATCCG	66
QY	61	CTGCGCCCCCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCCAGCCCGAG	120
Db	67	CTGCGCCCCCGCGCAAGAAAGTACAAAGCTGAAGCACTGTGTGGGCCAGCCCGAG	126
QY	121	CTGGAGGGCTTCGCGCTTGAAACCCCGGCTGTCTGAGACCGCCGAGGCTGCAAGCAGATC	180
Db	127	CTGGAGCGCTTCGCGCTTGAAACCCCGGCTGTCTGAGACCAAGCGAGGCTGCGCAGATC	186
QY	181	ATGAAGCAGCTGCAAGCCCGCCCTGCAAGACCGGCAACGAGAGCTGCGCAGCCTGTACAAC	240
Db	187	CTGGGCGCAGCTGCAGCCCAAGCCTGCAGACCGGCAGCGAGAGCTGCGCAGCCTGTACAAC	246
QY	241	ACCGTGGCCACCCCTGTACTGCGTGCAAGCCGCGCATGAGGTTCCGCGACCAAGAGGCC	300
Db	247	ACCGTGGCCACCCCTGTACTGCGTGCAAGCCGCGCATGAGGTTCAAGGACCAAGAGGCC	306
QY	301	CTGGAACAAGATCGAAGAGAGAGAGACAAGTCCCAAGACAAGACCAAGAGCCCAAGAG	360
Db	307	CTGGAAGAAGATCGAAGAGAGAGAGACAAGTCCAAAGAAAGGCCCAAGAGGCCCGCGCC	366
QY	361	GCCGACGGCA-----AGGTAGCCAGAACTACCCCATCTGTGCAGAACCTG	405
Db	367	GCCGCGCGCACCGGCAACAGCAGCCAGGTAGCCAGAACTACCCCATCTGTGCAGAACCTG	426
QY	406	CAGGGCCAGATGTGCACCAAGCCATCAGCCCGCACCCCTGAACCGCTGGTGAAAGTG	465
Db	427	CAGGGCCAGATGTGCACCAAGCCATCAGCCCGCACCCCTGAACCGCTGGTGAAAGTG	486
QY	466	ATCGAGAGAAAGCCTTCAGCCCGGAGGTGATCCCATGTTACCGCCCTGAGCGAGGC	525
Db	487	GTGAGAGAGAAAGCCTTCAGCCCGGAGGTGATCCCATGTTACCGCCCTGAGCGAGGC	546
QY	526	GCCACCCCCCAAGACTTGAAACAGATGTTGAACAACCTGGGCGGCAACAGGCGCGCATG	585
Db	547	GCCACCCCCCAAGACTTGAAACAGATGTTGAACAACCTGGGCGGCAACAGGCGCGCATG	606
QY	586	CAGATGCTGAAGGACACCATCAACGAGGAGCGCGCGAGTGGAACCGCTGCACCCCGTG	645
Db	607	CAGATGCTGAAGGAGACCATCAACGAGGAGCGCGCGAGTGGAACCGCTGCACCCCGTG	666
QY	646	CAGGCGGCGCCGCTGGCCCCCGGCGAGATGCGGACCCCCGCGGCAAGCATGCGCGG	705
Db	667	CAGGCGGCGCCGCTGGCCCCCGGCGAGATGCGGAGCCCCGCGGCAAGCATGCGCGG	726
QY	706	GCCACCAAGCAACCTTGAGAGAGATGCGCTGATGACCAAGCAACCCCCCGTGCCTG	765
Db	727	ACCAACCAAGCAACCTTGAGAGAGATGCGCTGATGACCAACCAACCCCCCATCCCGTG	786
QY	766	GCGCATCTACAAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGATGTACAGC	825
Db	787	GCGCATCTACAAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGATGTACAGC	846
QY	826	CCCGTGAGCATCTTGACATCGCCAGGGCCCAAGAGCCCTTCGCGCACTACGTGAC	885

```

Db      847 CCCACCAGCATCCTGGACATCCGCCAGGGGCCCAAGAGCCCTTCCGCACTACGTGAC 906
QY      886 CGCTTCTTCAAGACCTTGGCGCGCCGAGCAGGCCACCCAGAGCGTGAAGACTGATGACC 945
Db      907 CGCTTCTTCAAGACCTTGGCGCGCTGAGCAGGCCACCCAGAGCGTGAAGACTGATGACC 966
QY      946 GAGACCTTGTGTGTCAGAAACGCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGC 1005
Db      967 GAGACCTTGTGTGTCAGAAACGCCAACCCCGACTGCAAGACCATCTCGAGGCTCTCGGC 1026
QY      1006 CCGGCGGCCACCTTGAAGAGATGATGACCTGCTGCGAGGGCGTGGGGCGCCCGCGCAC 1065
Db      1027 CCGGCGGCCACCTTGAAGAGATGATGACCTGCTGCGAGGGCGTGGGGCGCCCGCGCAC 1086
QY      1066 AAGGCCCGCTGTGTCGGCGAGCGGATGAGCCAGG---CCAAACAGCGTGAACATCATGATG 1122
Db      1087 AAGGCCCGCTGTGTCGGCGAGCGGATGAGCCAGGATGAGCAACCCGCGCAACATCATGATG 1146
QY      1123 CAGAAAGCAACTTCAAGGGGCCCGCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAG 1182
Db      1147 CAGCGCGGCAACTTCCGCAACCGCGGAAGACCGTCAAGTCTTCAACTGCGGCAAGAG 1206
QY      1183 GGGCACAATGCGCAAGAACTGCGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGCAAG 1242
Db      1207 GGGCACAATGCGCAAGAACTGCGCGCGCGCGCGCGCAAGAGGGCTGTGCGCGCGCGC 1266
QY      1243 GAGGGCCACCAAGATGAAGAACTGACCGAGCGCGCGCAACTTCTGCGCAAGATCTGG 1302
Db      1267 GAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTTAAAGGAAGATCTGG 1326
QY      1303 CCCAGCCCAAGAGGGCGCGCGCGCAACTTCTGCAAGAACCGCAGCGAGCCCGCGCGCGC 1362
Db      1327 CTTCTCTCAAGAGGGAAGCGCGCAAGGAAATTTTCTTCAAGACAG-----A 1368
QY      1363 ACCGTGCGCCACCGCGCGCGCGCGCGCAAGCTTCCGCTTGAAGAGCAACCCCGCGCGC 1422
Db      1369 CCAGAGCCCAAGAGCGCGCGCGCAAGAGAGCTTCAAGTTTGGGAGAGAAAAACAATCTCC 1428
QY      1423 AAGCAGAGAGCCCAAGAGCGCGCGCTTACCGCGAGCGCGCTGACCGCGCTGCGCAGCGCTG 1482
Db      1429 TCTCAGAGAGCAGAGCGCGGATGACAGGAAGTCTTAACTTCCCTCAGATCACTC 1488
QY      1483 TTGCGCAGCGCGCGCGCTGAGCCAGTAA 1509
Db      1489 TTGGCAACGAGCCCTCTGTACAGTAA 1515

```

```

RESULT 8
US-09-475-515-78
; Sequence 78, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475, 515A
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 1865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```

OTHER INFORMATION: Description of Artificial Sequence: GPI  
US-09-475-515-78

Query Match 74.1%; Score 1117.8; DB 4; Length 1865;  
Best Local Similarity 85.1%; Pred. No. 1e-160;  
Matches 1299; Conservative 0; Mismatches 192; Indels 36; Gaps 3;

```

QY      1 ATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db      13 ATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72
QY      61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db      73 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 132
QY      121 CTGGAAGGCTTCCGCTGAAACCCCGCGCTGTGAGAACCGCGCGAGGCTGCGCAAGCATC 180
Db      133 CTGGAAGGCTTCCGCTGAAACCCCGCGCTGTGAGAACCGCGAGGCTGCGCGAGATC 192
QY      181 ATGAAGCAGCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db      193 CTGGGCGCAAGCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252
QY      241 ACCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db      253 ACCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
QY      301 CTGGAACAAGATGAGAGAGAGCAAGAACAGTCCAGCAGAAAGACCCAGAGGCCCAAGAGAG 360
Db      313 CTGGAAGAAGATGAGAGAGAGCAAGAACAGTCCAAAGAAAGGCCAGAGCGCGCGCGCGC 372
QY      361 GCCGACGGCA-----AGGTGAGCCAGAACTAACCCCATCTGCGAAGACTTG 405
Db      373 GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
QY      406 CAGGGCCAAATGTGTGCAACCAAGGCCATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 465
Db      433 CAGGGCCAAATGTGTGCAACCAAGGCCATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492
QY      466 ATCGAGGAGAAAGGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 525
Db      493 GTGAGGAGAAAGGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 552
QY      526 GCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 585
Db      553 GCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
QY      586 CAGATGCTGAAGAGCAACCATCAACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
Db      613 CAGATGCTGAAGAGAGCAACCATCAACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 672
QY      646 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705
Db      673 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732
QY      706 GCCAAGCAGCAACCTTGAAGAGAGCAAGTGGCTGTGATGACCAAGCAACCCCGTGGCGCGTG 765
Db      733 ACCAAGCAGCAACCTTGAAGAGAGCAAGTGGCTGTGATGACCAAGCAACCCCGTGGCGCGTG 792
QY      766 GGGCAGCATCTCAAGCGGTGATCATCTTGGGCTTGAACAAGATGTGCGGATGTACAGC 825
Db      793 GGGCAGATCTCAAGCGGTGATCATCTTGGGCTTGAACAAGATGTGCGGATGTACAGC 852
QY      826 CCGGTGAGCATCTTGAAGATCCGCGCAGGGGCCCAAGAGGCCCTTCCGCGACTACGTGAGAC 885
Db      853 CCCAAGCAGCATCTTGAAGATCCGCGCAGGGGCCCAAGAGGCCCTTCCGCGACTACGTGAGAC 912
QY      886 CGCTTCTTCAAGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945
Db      913 CGCTTCTTCAAGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 972
QY      946 GAGACCTGTGTGTGAGAGCGCGCAACCCGACTGCAAGACCATCTGCGCGCTCTCGGC 1005

```

Db 973 GAGACCTGCTGTGTCAGAACGCCAACCCCGACTGCAAGACCATCTGAAGGCTCTCGGC 1032  
QY 1006 CCCGCGCCACCCCTGAGAGATGATGACCGCTGCCAGGGCGTGGCGGCCCGCGCAC 1065  
Db 1033 CCCGCGCCACCCCTGAGAGATGATGACCGCTGCCAGGGCGTGGCGGCCCGCGCAC 1092  
QY 1066 AAGGCCCGCTGTGGCGGAGCGATGAGCCAGG---CCAACAGCGTGAACATCATGATG 1122  
Db 1093 AAGCCCCGCTGTGGCGGAGCGATGAGCCAGGTGACGAACCCGGCCACCATCATGATG 1152  
QY 1123 CAGAAAGCAACTTCAAGGGCCCCCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAG 1182  
Db 1153 CAGCGCGCAACTTCCGCAACCAAGCGGAAGACCGTCAAGTCTTCAACTGCGGCAAGAG 1212  
QY 1183 GGGCACAATGCGCAAGACTGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGCAAG 1242  
Db 1213 GGGCACAATGCGCAAGACTGCGCGCGCGCGCGCAAGAGGGCTGTGCGCGCTGGCGCGC 1272  
QY 1243 GAGGCGCAACGATGAAGGACTGCAACCGGCGCGCAACTTCTGGGCAAGATCTGG 1302  
Db 1273 GAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTTAAAGGAAGATCTGG 1332  
QY 1303 CCCAGCCACAAGGGCGCGCGCGCAACTTCTGCAAGAACCGGAGCGAGCGCGCGCGCC 1362  
Db 1333 CTTTCTTCAAGGGAAGGCGCAGGGAATTTTCTTCAAGAGAG-----A 1374  
QY 1363 ACCGTGCCCCACCGCGCGCGCGCGCGCGCTTCCGCTTGAAGAGACCAACCGCGCGCC 1422  
Db 1375 CCAGAGCCAAACAGCGCCCAACAGAGAGGCTTCAAGTTTGGGAGAGAAAACCACTGCC 1434  
QY 1423 AAGCAGAGCCCAAGGACCGGAGCGCTTACCGGAGCGCGCTGACCGCGCTGCGCAGCTG 1482  
Db 1435 TCTCAGAAAGCAGAGCGGATGACAAAGAACTGATCTTAACTTCCCTCAGATCACTC 1494  
QY 1483 TTGCGCAGCGCGCGCGCTGAGCCAGTAA 1509  
Db 1495 TTGCGCAACGACCCCTGCTCAGAGTAA 1521

RESULT 9

US-09-475-515-79  
; Sequence 79, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 79  
; LENGTH: 1865  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GP2  
US-09-475-515-79

Query Match 74.1%; Score 1117.8; DB 4; Length 1865;  
Best Local Similarity 85.1%; Pred. No. 1e-160;  
Matches 1299; Conservative 0; Mismatches 192; Indels 36; Gaps 3;

QY 1 ATGGCGCGCGCGCCAGCATCTGCGCGCGGAGAGCTGACAAAGTGGAGAGAGATCCGC 60

Db 13 ATGGCGCGCGCGCCAGCGTGTGAGCGCGCGGAGCTGGAACAAGTGGAGAGATCCGC 72  
QY 61 CTGCGCGCGCGCGCAAGAGCACTATGCTGAAGCACTGTGTGGCCAGCGCGAG 120  
Db 73 CTGCGCGCGCGCGCAAGAGCACTATGCTGAAGCACTGTGTGGCCAGCGCGAG 132  
QY 121 CTGAGGGCTTGCCTGAACCCGCGCTGTGAGACCGCGGAGGCTGCAAGCAGATC 180  
Db 133 CTGAGCGCTTGCCTGAACCCGCGCTGTGAGACCGCGGAGGCTGCGCGAGATC 192  
QY 181 ATGAAGCAGTGAAGCGCGCGCTGAGACCGGAGAGAGCTGCGCAGCTGTACAAC 240  
Db 193 CTGGCCAGCTGCAAGCCAGCTGAGACCGGAGAGAGAGCTGCGCAGCTGTACAAC 252  
QY 241 ACCGTGGCCACCTGTACTGCGTGAAGCGCGGCAATCGAGTCCGCGACCAAGAGGCC 300  
Db 253 ACCGTGGCCACCTGTACTGCGTGAAGCGCGGCAATCGAGTCCGCGACCAAGAGGCC 312  
QY 301 CTGGAACAAGATGAGAGAGAGCAAGTCCAGCAAGAACCAAGCAAGCGCAAGAG 360  
Db 313 CTGGAAGAAGATGAGAGAGAGCAAGTCCAGCAAGAACCAAGCGCGCGCGCC 372  
QY 361 GCCGACGGCA-----AGGTGAGCCGAAGTACCCTATCGTGAAGACCTG 405  
Db 373 GCCGCGCGGCAACCGGCAACAGCAGCGTGAAGCCAGACTACCCCATCGTGAAGACCTG 432  
QY 406 CAGGCGCAGATGTGTCACCAAGGCAATCAGCCCGCGCAACCTGAACGCTGGTGAAGTG 465  
Db 433 CAGGCGCAGATGTGTCACCAAGGCAATCAGCCCGCGCAACCTGAACGCTGGTGAAGTG 492  
QY 466 ATCAGAGAGAGGCTTCAAGCGCGCGAGGTGATCCCATGTTCAAGCGCGCTGAGCGAGGC 525  
Db 493 GTGAGAGAGAGGCTTCAAGCGCGCGAGGTGATCCCATGTTCAAGCGCGCTGAGCGAGGC 552  
QY 526 GCCACCGCGCGAGCACTGAACAGATGTTGAACACCGTGGCGGCAACGAGCGCGCATG 585  
Db 553 GCCACCGCGCGAGCACTGAACAGATGTTGAACACCGTGGCGGCAACGAGCGCGCATG 612  
QY 586 CAGATGCTGAAGAGCAACCATCAACGAGAGGCGCGCGAGTGGACCGCTGCACCGCGTG 645  
Db 613 CAGATGCTGAAGAGAGCAACCATCAACGAGAGGCGCGCGAGTGGACCGCTGCACCGCGTG 672  
QY 646 CAGGCGCGCGCGCGCGCGCGCGCGCGAGTGGCGGACCGCGCGCGCGCGCGCGCGCG 705  
Db 673 CAGGCGCGCGCGCGCGCGCGCGCGCGAGTGGCGGACCGCGCGCGCGCGCGCGCGCG 732  
QY 706 GCCACCGAGCACTGAGAGAGAGTGGCTGATGACCAAGCAACCGCGCGCGCGTG 765  
Db 733 ACCACCGAGCACTGAGAGAGAGTGGCTGATGACCAAGCAACCGCGCGCGCGTG 792  
QY 766 GGGGACATTTACAGCGGCTGATCATCTGGGCGCTGAACAAAGATGCTGGGATGACAGC 825  
Db 793 GGGGACATTTACAGCGGCTGATCATCTGGGCGCTGAACAAAGATGCTGGGATGACAGC 852  
QY 826 CCGGTGAGCATCTGAGCATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885  
Db 853 CCGGTGAGCATCTGAGCATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912  
QY 886 CGCTTCTTCAAGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945  
Db 913 CGCTTCTTCAAGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 972  
QY 946 GAGACCTGTGTGTCAGAACGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1005  
Db 973 GAGACCTGTGTGTCAGAACGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1032  
QY 1006 CCCGCGCGCCACCTGAGAGAGATGATGACCGCTGCCAGGGCGTGGCGCGCGCGCGCAC 1065  
Db 1033 CCCGCGCGCCACCTGAGAGAGATGATGACCGCTGCCAGGGCGTGGCGCGCGCGCGCAC 1092  
QY 1066 AAGGCCCGCTGTGGCGGAGCGATGAGCCAGG---CCAACAGCGTGAACATCATGATG 1122

Db 1093 AAGCCCCGCTGCTGCGCGAGGCGATGAGCCAGGTGACGAACCCGGCGACCATCATGATG 1152  
QY 1123 CAGAGAGCACTTCAAGGGCCCCCGGCGCAACGTCAAAGTCTTCAACTGGCGCAAGAG 1182  
Db 1153 CAGCGCGCACTTCCGCAACAGCGGAAGACCGTCAAGTCTTCAACTGGCGCAAGAG 1212  
QY 1183 GGGCACATCGCCAAAGACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAG 1242  
Db 1213 GGGCACACCGCCAGAACTGCGCGCCCCCGCAAGAGGGCTGTGCGCGCGCGC 1272  
QY 1243 GAGGCGCAACAGATGAAGACTGCAACCGAGCGCGCAACTTCTGGGCAAGATCTGG 1302  
Db 1273 GAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTTAATTTTAAAGGAAGATCTGG 1332  
QY 1303 CCCAGCAACAAGGGCGCGCGCAACTTCTGCAAGAACCGAGCGAGCCGCGCGCGCC 1362  
Db 1333 CCTTCTTAACAAGGGAAGGCGAGGAATTTTCTTCAAGACAG-----A 1374  
QY 1363 ACCGTGCCCCACCGCGCGCGCGCGCAAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCC 1422  
Db 1375 CCAGAGCCCAACAGCGCGCGCAAGAGAGCTTCAAGTTTGGGAGAGAGAAAACAATCTCC 1434  
QY 1423 AAGCAGAGAGCCCAAGAGCGCGAGCGCTTACCGGAGCGCGCTGACCGCGCGAGCGCTG 1482  
Db 1435 TCTCAGAGAGCAGAGAGCGGATAGACAAGAACTGTATCTTTAACTTCCCTCAGATCACTC 1494  
QY 1483 TTGCGCAGCGCGCGCGCGCGCGCACTTAA 1509  
Db 1495 TTGGCAACGACCGCGCTGTCACAGTAA 1521

## RESULT 10

US-09-475-515-6  
/ Sequence 6, Application US/09475515A  
/ Patent No. 6602705  
/ GENERAL INFORMATION:  
/ APPLICANT: BARNETT, Susan  
/ APPLICANT: ZUR MEGEDE, Jan  
/ APPLICANT: SRIVASTAVA, Indresh  
/ APPLICANT: LIAN, Ying  
/ APPLICANT: HARTOG, Karin  
/ APPLICANT: LIU, Hong  
/ APPLICANT: GREER, Catherine  
/ APPLICANT: SREBY, Mark  
/ APPLICANT: WALKER, Christopher  
/ TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
/ TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
/ FILE REFERENCE: 1621.002  
/ CURRENT APPLICATION NUMBER: US/09/475,515A  
/ CURRENT FILING DATE: 1999-12-30  
/ NUMBER OF SEQ ID NOS: 90  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 6  
/ LENGTH: 4319  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic  
/ OTHER INFORMATION: HIV-gag-polymerase  
US-09-475-515-6

Query Match 74.1%; Score 1117.8; DB 4; Length 4319;  
Best Local Similarity 85.1%; Pred. No. 1e-160;  
Matches 129; Conservative 0; Mismatches 192; Indels 36; Gaps 3;

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGGAGAGAGTGAAGTGGAGAGATCCGC 60  
Db 7 ATGGGCGCGCGCGCAGCGTGTGAGCGCGCGGAGAGTGAAGTGGAGAGATCCGC 66  
QY 61 CTGCGCGCGCGCGCAAGAGCACTAATGCTGAAGACACTGTGTGGCGCGCGAG 120  
Db 67 CTGCGCGCGCGCGCAAGAGAGTGAAGTGAAGCACTGTGTGGCGCGCGAG 126

QY 121 CTGAGGGCTTCCCTTGAACCCCGCGCTGTGAGAGCCGCGAGGGCTGCAAGCAGATC 180  
Db 127 CTGAGCGCTTCCCTTGAACCCCGCGCTGTGAGAGCCGCGAGGGCTGCGCGCAGATC 186  
QY 181 ATGAAGCAGCTGAGCCCGCGCTGAGACCGCGCAACGAGAGCTGCGAGCTGTACAAC 240  
Db 187 CTGGGCCAGCTGAGCCCGCGCTGAGACCGCGCAACGAGAGAGCTGCGAGCTGTACAAC 246  
QY 241 ACCGTGGCACCTGTACTGTGTCAGCGCGCGCATGAGTCCGCGACCAAGAGGCC 300  
Db 247 ACCGTGGCACCTGTACTGTGTCAGCGCGCATGAGTCCGCGACCAAGAGAGGCC 306  
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTCCAGCAGAAAGACCAGAGGCCAAGAG 360  
Db 307 CTGGAAGAAGATCGAGAGAGAGCAACAAGTCCAAAGAAAGGCCAGAGGCCCGCGCC 366  
QY 361 GCCGACGGCA-----AGGTAGCCAGAACTAACCCCATCTGTGCAAACTG 405  
Db 367 GCCGCGGCAACCGGCAACAGACAGCAGGTGAGCCAGAACTAACCCCATCTGTGCAAACTG 426  
QY 406 CAGGGCCAGATGTGACACAGGCGCATCAGCCCGCGCACCTGAAACGCTGGTGAAGTG 465  
Db 427 CAGGGCCAGATGTGACACAGGCGCATCAGCCCGCGCACCTGAAACGCTGGTGAAGTG 486  
QY 466 ATCGAGAGAAAGCTTCAAGCGCGAGGTGATCCCATGTTCAACGCGCTGACGAGGCG 525  
Db 487 GTGAGAGAAAGCTTCAAGCGCGAGGTGATCCCATGTTCAACGCGCTGACGAGGCG 546  
QY 526 GCCACCCCGCGAGACCTGAACAAGTGTGAACAACCGTGGCGCGCACAGCGCGCATG 585  
Db 547 GCCACCCCGCGAGACCTGAACAAGTGTGAACAACCGTGGCGCGCACAGCGCGCATG 606  
QY 586 CAGATGCTGAAGAGACCATGAACAAGAGAGCGCGCGAGTGGAGCCGCTGCAACCCGCTG 645  
Db 607 CAGATGCTGAAGAGACCATGAACAAGAGAGCGCGCGAGTGGAGCCGCTGCAACCCGCTG 666  
QY 646 CAGCGCGCGCGGTGGCGCGCGCGAGATGCGGAGCCCGCGCGAGCAGATCGCGGC 705  
Db 667 CAGCGCGCGCGCATGCGCGCGCGCGAGATGCGGAGCCCGCGCGAGCAGATCGCGGC 726  
QY 706 GCCACAGACCTTGAAGAGCAGATGCGCTGATGACCAACCCCGCTGCGCGCTG 765  
Db 727 ACCACAGACCTTGAAGAGCAGATGCGCTGATGACCAACCCCGCTGCGCGCTG 786  
QY 766 GCGGACATCTAAGCGGTGATCATCTGCGCTGAACAAGATCGTGCAGATGACAGC 825  
Db 787 GCGGAGATCTAAGCGGTGATCATCTGCGCTGAACAAGATCGTGCAGATGACAGC 846  
QY 826 CCGGTGAGCATCTGGAATCCGCGAGGCGCGCAAGAGCCCTTCCGCGACTAGTGAGC 885  
Db 847 CCGGAGATCTGGAATCCGCGAGGCGCGCAAGAGCCCTTCCGCGACTAGTGAGC 906  
QY 886 CGCTTCTTAAGACCTGCGCGCGCGAGCAGCGCACCGCAGAGTGAAGAGTGAAGTGAAC 945  
Db 907 CGCTTCTTAAGACCTGCGCGCGCTGAGCAGCGCGCACCGCAGAGTGAAGAGTGAAGTGAAC 966  
QY 946 GAGACCTGTGTGAGAGAGCGCAACCGCGACTGCAAGACCATCTGCGCGCTGCGGC 1005  
Db 967 GAGACCTGTGTGAGAGAGCGCAACCGCGACTGCAAGACCATCTGCGCGCTGCGGC 1026  
QY 1006 CCGGCGCGCACCTGAGAGAGTGAACCGCTGCGAGGCGGTGGCGCGCGCGCGCAC 1065  
Db 1027 CCGGCGCGCACCTGAGAGAGTGAACCGCTGCGAGGCGGTGGCGCGCGCGCGCAC 1086  
QY 1066 AAGGCGCGGTGTGCGCGAGCGATGAGCGAG---CCAACAGCGTGAACATCATGATG 1122  
Db 1087 AAGGCGCGGTGTGCGCGAGCGATGAGCGAGCGAGCGCGCGCATCATGATG 1146  
QY 1123 CAGAAGCAACTTCAAGGGCGCGCGCAAGTGAAGTGTCTTCAACTGGCGCAAGAG 1182  
Db 1147 CAGCGCGCAACTTCCGCAACAGCGGAAGACGTCAAGTGTCTTCAACTGGCGCAAGAG 1206  
QY 1183 GGCACATCGCCAAAGAGTGTGCGCGCGCGCGCAAGAGAGGCTGTGGAAGTGGCGCAAG 1242

Db 1207 GGCCACCCGCCAGGAAGTCCCGCGCCCCCGCAGAAAGGCGTGCCTGGCGCTGCGCGCCG 1266

QY 1243 GAGGCGCACCGATGAAAGACTGCACCGAGCGCCAGGCCACTTCTTGAGCAAGATCTGG 1302

Db 1267 GAAGGACACCAATGAAAGATTGGCACTGAGAGACAGCTTAATTTTAAAGGAGATCTGG 1326

QY 1303 CCGAGCCACAAGGCGCGCCCGCGCACTTCTGCAGAACCGCAGCGAGCCCGCGCCCC 1362

Db 1327 CCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTACAGCAG-----A 1368

QY 1363 ACCGTGCCCAACCGCCCCCCCCCGCCAGAGCTTCCGCTTGCAGAGACCACCCCCGCCCC 1422

Db 1369 CCAGAGCCCAACAGCCCCCACCAGAAAGAGCTTCAAGTTTGGGGAGGAGAAAAACAATCCC 1428

QY 1423 AAGCAGGAGCCCAAGGACCGCGAGCCCTACCGCGAGCCCCCTGACCGGCCCTGGGCAAGCTG 1482

Db 1429 TCTCAGAAAGCAGGAGCCGATTAGACAAAGAACTGTATCTTTAACTTCCCTCAGATCACTC 1488

QY 1483 TTGGCAGCGCGCCCCCTGAGCCAGTAA 1509

Db 1489 TTGGCAACGACCCCTCGTCACAGTAA 1515

## RESULT 11

```

US-09-475-515-9
; Sequence 9, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karln
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1268
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag
; OTHER INFORMATION: common region
US-09-475-515-9

```

Query Match	69.5%;	Score 1049.2;	DB 4;	Length 1268;
Best Local Similarity	90.8%;	Pred. No. 2.4e-150;		
Matches 1146; Conservative	0;	Mismatches 98;	Indels 18;	Gaps 2;

OY	1	A T G G C G C C C G C G C C A G C A T C C T G C G C G G C G A G A A G C T T G G A C A A G T G G A G A A G A T C C G C	60
Db	7	A T G G C G C C C G C G C C A G C C T G C T G A G C G G C G G C G A G C T G G A C A A G T G G A G A A G A T C C G C	66
OY	61	C T G C G C C C C C G G C G G C A A G A G C A C T A C A T G C T G A A G C A C C T G T G T G G C C A G C C C G C A G	120
Db	67	C T G C G C C C C C G G C G G C A A G A G A G T A C A A G C T G A A G C A C A T C G T G T G G C C A G C C G C A G	126
OY	121	C T G A G G G C T T C G C C C T G A A C C C C G G C C T G C T G A G A C C G C G A G G G C T G C A A G C A G A T C	180
Db	127	C T G A G C G C T T C G C C G T G A A C C C C G G C C T G C T G A G A C A C G A G G G T G C C G C A G A T C	186
OY	181	A T G A A G C A G C T G C A G C C C G C C C T G C A G A C C G G C A C C G A C G A G A G C T G C G C A G C C T G T A C A C	240
Db	187	C T G G C C A G C T G C A G C C C A G C C T G C A G A C C G G C A G C G A G A G C T G C G A G C C T G T A C A C	246

QY	241	ACCGTGGCCACCCTGTA	CTGCGTGCA	CGCCGGCAT	CGAGGTCCGCGACA	CCCAAGAGGCC	300
Db	247	ACCGTGGCCACCCTGT	ACTGCTGCA	CCAGCGCAT	CGACGTCAAGGACA	CCCAAGAGGCC	306
QY	301	CTGGAACAAGATCGA	GAGGAGCA	GAAACAAGTCC	AGCAAGAACCC	AGAGGCCAAGAG	360
Db	307	CTGGAAGAAGATCGA	GAGGAGCA	GAAAGTCCA	AGAAAGGCCCA	AGAGGCCCGGCC	366
QY	361	GCCGACGGCA-----	AGGTAGCCAGAA	CTTACCCCAT	CGTGCAGAA	CTG	405
Db	367	GCCGACGGCA	CCGCGCAACA	GACGACGAGT	AGCCAGAACT	TACCCCATCGTGCAGAAC	426
QY	406	CAGGGCCAGATGGT	GCACCA	GGCCATCAG	CCCCCGCAC	CCCTGAAAGGTG	465
Db	427	CAGGGCCAGATGGT	GCACCA	GGCCATCAG	CCCCCGCAC	CCCTGAAAGGTG	486
QY	466	ATCGAGGAGAA	GGCCTT	CAGCCCCGAG	GTGATCCCAT	GTTCACCGCCCTGAGCGAGGC	525
Db	487	GTGAGGAGAA	GGCCTT	CAGCCCCGAG	GTGATCCCAT	GTTCACCGCCCTGAGCGAGGC	546
QY	526	GCCACCCCCC	CAGACCT	GAAACA	CGATGTTGA	CA	585
Db	547	GCCACCCCCC	CAGACCT	GAAACA	CGATGTTGA	CA	606
QY	586	CAGATGCTGA	AGGACA	CCATCA	CGAGAGGCGCC	CGAGTGGGACCGCTGCACCCCGTG	645
Db	607	CAGATGCTGA	AGGAGA	CCATCA	CGAGAGGCGCC	CGAGTGGGACCGCTGCACCCCGTG	666
QY	646	CAGGCGGGCCCCG	TGGCCCCCGG	CCAGATGCGCG	CA	CCCCCGGCGAGCGACATCGCCGGC	705
Db	667	CAGGCGGGCCCCC	ATCGCCCCCG	CCAGATGCGCG	CA	CCCCCGGCGAGCGACATCGCCGGC	726
QY	706	GCCACCA	GCA	CCCTG	CAGAGAGATCG	CGCTGATGACCA	765
Db	727	ACCA	CCAGCA	CCCTG	CAGAGAGATCG	CGCTGATGACCA	786
QY	766	GGCGACAT	CTACA	AGCGGTGAT	CTCTGGGCT	GAAACAAGATCGTGCAGATGTACAGC	825
Db	787	GGGAGAT	CTACA	AGCGGTGAT	CTCTGGGCT	GAAACAAGATCGTGCAGATGTACAGC	846
QY	826	CCCGTGAGCAT	CTGGA	CATCGCC	CAGGCGCCCA	AGAGCCCTTCGCGACTACGTGAC	885
Db	847	CCCA	CCAGCAT	CTGGA	CATCGCC	CAGGCGCCCA	906
QY	886	CGCTTCTT	CAAGACC	CTGCGGCG	AGAGCCCA	AGAGCCCTTCGCGACTACGTGAC	945
Db	907	CGCTTCTT	CAAGACC	CTGCGGCG	AGAGCCCA	AGAGCCCTTCGCGACTACGTGAC	966
QY	946	GAGAC	CCCTG	GTGAGAA	CGCCAA	CCCGACTGCAAGACATCTCGCGCTCTCGC	1005
Db	967	GAGAC	CCCTG	GTGAGAA	CGCCAA	CCCGACTGCAAGACATCTCGCGCTCTCGC	1026
QY	1006	CCCGGCGCA	CCCTG	AGAGAGAT	GTACCG	CTGCAAGGCGTGGCGGCCCGGCCAC	1065
Db	1027	CCCGGCGCA	CCCTG	AGAGAGAT	GTACCG	CTGCAAGGCGTGGCGGCCCGGCCAC	1086
QY	1066	AAGG	CCGCGT	GCTGGCCG	AGGCGAT	GAGCAGG--CCA	1122
Db	1087	AAGG	CCGCGT	GCTGGCCG	AGGCGAT	GAGCAGG--CCA	1146
QY	1123	CAGAA	GAGCA	ATTCA	AGGCCCCCG	CGCAAGTCAAGTGTCTTCA	1182
Db	1147	CAGCG	CGGCA	ATTCCG	CAACCA	CGGAAGCCGTCAAGTGTCTTCA	1206
QY	1183	GGCCACAT	CGCCAA	GAACTG	CGCGCC	CCCGCAAGAGGCTGTGAAGTGGCGCAAG	1242
Db	1207	GGCCACAT	CGCCAA	GAACTG	CGCGCC	CCCGCAAGAGGCTGTGAAGTGGCGCGC	1266
QY	1243	GA	1244				
Db	1267	GA	1268				



```
RESULT 13
US-09-974-702-1
; Sequence 1, Application US/09974702
; Patent No. 6696291
; GENERAL INFORMATION:
;   APPLICANT: Shiver, John W.
;             Davies, Mary Ellen
;             Freed, Daniel C.
;             Liu, Margaret A.
;             Perry, Helen C.
; TITLE OF INVENTION: Synthetic HIV Gag Genes
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Merck & Co., Inc.
;   STREET: P.O. Box 2000, 126 E. Lincoln Ave.
;   CITY: Rahway
;   STATE: NJ
;   COUNTRY: USA
;   ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/974,702
;   FILING DATE: 09-Oct-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 09/017,981
;     FILING DATE: 03-FEB-1998
;     APPLICATION NUMBER: US60/037,854
;     FILING DATE: 07-FEB-1997
;     APPLICATION NUMBER: GB9705040.5
;     FILING DATE: 12-MAR-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Hand, J. Mark
;       REGISTRATION NUMBER: 36,545
;       REFERENCE/DOCKET NUMBER: 19730
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 732-594-3905
;       TELEFAX: 732-594-4720
;     TELEX: <Unknown>
;     INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH: 1532 base pairs
;         TYPE: nucleic acid
;         STRANDEDNESS: double
;         TOPOLOGY: both
;         MOLECULE TYPE: cDNA
;         HYPOTHETICAL: NO
;         ANTI-SENSE: NO
;         SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-974-702-1

Query Match      64.4%; Score 971.4; DB 4; Length 1532;
Best Local Similarity 79.6%; Pred. No. 1.4e-138;
Matches 1210; Conservative 0; Mismatches 281; Indels 30; Gaps 4;
```

```
Db      190 CTGGCCAGCTCCAGCCCTCCCTGCAACAGGCTTGAGGAGCTGAGTCCCTGTACAAC 249
Qy      241 ACCGTGGCCACCCTGTACTGCTGTGACGCGGCATCGAGTCCGCGACCAAGAGGCC 300
Db      250 ACAGTGGCTACCTGTACTGTGTGACCAAGAAATGTATGTGAAGACACCAAGAGGCC 309
Qy      301 CTGCAAGATCGAGAGGAGAGCAAGTCCAGCAAGAGAGCCAGCAGGC----- 353
Db      310 CTGGAAGAATGAGAGAGAGCAAGTCAAGAAAGAGGCCAGCAGCTGTGCT 369
Qy      354 --CAAGAGGCCGCAAGGTGAGCCAGAACTACCCATCGTGAGAACTGCAAGGCC 411
Db      370 GGCAAGGCAACTCCAGCCAGGTGTCAGAACTACCCATTTGTGAGAACTCCAGGCC 429
Qy      412 CAGATGTGACAGAGCCATCAGCCCCCGACCTGAACGCTGGGTGAAGTGTAG 471
Db      430 CAGATGTGACAGAGCCATCTCCCCCGGACCTGAATGCTGGGTGAAGTGTAG 489
Qy      472 GAGAGGCTTCAAGCCCCGAGGTATCCCATGTTCAACGCCCTGAGCGGCGCAC 531
Db      490 GAGAGGCTTCTCCCTGAGGTATCCCATGTTCTTGCCCTGTCTGAGGCTGCAAC 549
Qy      532 CCCCAAGACCTGAACAGATGTTGAACACCGTGGCGGCCACAGCGCCCATGAGATG 591
Db      550 CCCCAAGACCTGAACACCATGCTGAACAGTGGGGGCCATCAGGCTGCATGAGATG 609
Qy      592 CTGAAGGACACCATCAAGAGAGCGCGCGAGTGGGACCGCTGCAACCCGTGAGGCC 651
Db      610 CTGAAGGAGACCATCAATGAGAGGCTGCTGAGTGGGACAGGCTGCATCTGTGACGCT 669
Qy      652 GGGCCCTGTGGCCCCCGGCGAGATGGCGGACCCCGCGGCGAGCATCGCGCGCAC 711
Db      670 GGGCCCTGTGGCCCCCGGCGAGATGGCGGACCCCGCGGCGAGCATCGCGCGCAC 729
Qy      712 AGCACCTTGACAGGACAGATCGCTGTGATGACCAACACCCCGCTGCGGCGAC 771
Db      730 TCCACCTTCAAGGACAGATTTGGCTGATGACCAACACCCCGCTGCGGCGAA 789
Qy      772 ATCTAAGCGGTGATCATCTGTGGGCTGAACAAGATGTCGCGATGTACAGCCCGTG 831
Db      790 ATCTAAGAGGTGATCATCTGTGGGCTGAACAAGATTTGTGAGATGTACTCCCCACC 849
Qy      832 AGCATCTTGACATCCGCCAGGCGCCCAAGAGGCTTCCGCGACTACGTGACCGCTTC 891
Db      850 TCCATCTTGACATCAGGAGGCGCCCAAGAGGCTTCAAGGACTATGTGACAGGTTTC 909
Qy      892 TTCAAGACCTTGCGCGCGCGGAGAGGCGCAACCAAGACGTGAAGAACTGGATGACCGAAC 951
Db      910 TACAAGACCTTGAGGCGCTGAGCAGGCTCTCCAGAGAGTGAAGAACTGGATGACAGAAC 969
Qy      952 CTGCTGTGCAAGACGCCAACCCGCACTGCAAGACCATCTGCGGCTCTCGGCCCGGC 1011
Db      970 CTGCTGTGCAAGATGCCAACCTGACTGCAAGACCATCTGAAAGGCCCTTGCGGCTGCT 1029
Qy      1012 GCCACCTTGAGAGATGATGACCGCTGCGCAGGCGTGGCGGCGCGGCCACAAAGGCC 1071
Db      1030 GCCACCTTGAGAGATGATGACCGCTGCGCAGGCGTGGCGGCGCGGCCACAAAGGCC 1089
Qy      1072 CGCGTGTGCGCGGCGGATGAGCGAG--CAACAGCGTGAACATCATGATGAGAAAG 1128
Db      1090 AGGGTGTGCTGAGGCGCATGTCCAGGTGACCACTCCGCGACATCATGATGAGAGG 1149
Qy      1129 AGCAACTTCAAGGCGCGCGCGCAAGTCAAGTCTTCAACTGCGCGCAAGAGGCCAC 1188
Db      1150 GGCAACTTCAAGAACTCAAGAGAAAGTCAAGTCTTCAACTGCGCGCAAGAGGCCAC 1209
Qy      1189 ATGCCCAAGAACTGCGCGCGCGCGCAAGAGGCTGTGAAAGTGGCGCAAGAGGGC 1248
Db      1210 ATTGCCAAGAACTGTAGGGCGCGCGCAAGAGAGGCTGTGAAAGTGTGGCAAGAGGGC 1269
Qy      1249 CACCATGAAGACTGCAACCGAGCGCGAGGCCAACTTCTGGGCAAGATCTGGCGCAGC 1308
```

Db	1270	CACCAGATGAAGGACTGCAMTGAAGGCAAGGCCAACTTCCTGGGCAAAATCTGGCCCTCC	1329
QY	1309	CACAAAGGGCCGCCCCGGGCAACTTCTCTGCAGAAACCGCAGCGAGCCCGGCCCAACCGTG	1368
Db	1330	CACAAAGGGCAGGCTGTGCMACTTCTCCAGTCCAGGCTGAGCCCAAGCCCTCTCCGAG	1389
QY	1369	CCCAACGGCCCCCGCCGAGAGCTTCGCTTGAGGAGAACCAACCCCGCCCAAGCAG	1428
Db	1390	GAGTCCTTC-----AGGTTGGGAGAGAAAGACCACCCCCAGCAAGAACAG	1437
QY	1429	GAGCCCAAGGACCGCGAGCCCTTAACCGCAGGCCCTGACCGCCCTGCGACGCTGTTGAGC	1488
Db	1438	GAGCCCAATTGACAAAGGAGCTGAC-----CCCTGGCCTCCCTGAGGTCCCTGTTTGGC	1491
QY	1489	AGCGGGCCCCCTGAGCCAGTAA	1509
Db	1492	AACGACCCCTCTCTCCAGTAA	1512

RESULT 14  
US-09-818-443-1  
; Sequence 1, Application US/09818443

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	64.4%	Score 971.4;	DB 4;	Length 1532;
Best Local Similarity	79.6%;	Pred. No. 1.4e-138;		
Matches 1210; Conservative	0;	Mismatches 281;	Indels 30;	Gaps 4;

QY 1 ATGGCGCCCCGCGCCAGCATCCTGCGCGGCGAGAACTGGACAAGTGGAGAATCCGC 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 10 ATGGGTGTAGGGCTTTCTGTCTGTCTGTGTGTAGACTGGACAACTGGAGAAGATCAGG 69  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 61 CTGCGCCCCGGCGGCAAGAACATACTGCTGAAGCACCTGTGTGTGGCCAGCCGGAG 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 70 CTGAGGCTGTGTGCAAGAAGAACTAACAGCTTAAGCACATTGTGTGGCTCCAGGAG 129  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 121 CTGAGGGCTTCGCCCTGAACCCCGGCTGTCTGAGACCGCCGAGGCTGCAAGCATC 180  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 130 CTGGAAGGTTTTGTGTGTGAACCCCTGGCTGTCTGAGACCTCTAAGGGGTGCAGGCAGATC 189  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 181 ATGAAGCAGCTGCAGCCCGCCCTGCAGACC GGACCGAGAGCTGCAGCCTGTACAAC 240  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 190 CTGGCCAGCTCCAGCCCTCCCTTGCAAACAGGCTCTGAGAGGTGAGGTCCCTGTACAAC 249  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 241 ACCGTGGCACCCCTGTACTGTGTGCAAGCCGGCATTCAGGTCCGCGACACCAAGAGGCC 300  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db	250	ACAGTGGCTACCTGTACTGTGTGCACAGAAAGATTGATGTAAGGACACCAAGAGAGCC	309
QY	301	CTGGACAGATCGAGGAGGAGCAGAACAAGTCCACAGAAAGACCAGCAGGC-----	353
Db	310	CTGGAGAAAGATTGAGGAGGAGCAGAACAAGTCCAAAGAAAGAGGCCAGCAGCTGTGCT	369
QY	354	--CAAGGAGGCCGACGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAGAACTTGACGGGC	411
Db	370	GGCACAGGCCAATCTCCAGCCAGGTGTCTCCAGAACTACCCCATTTGTGCAGAACTCCAGGGC	429
QY	412	CAGATGTTGCACCAAGGCCATCAGCCCCCGCACCCCTGAAACGCTGGGTGAAGGTATCGAG	471
Db	430	CAGATGTTGCACCAAGGCCATCTCCCCCGGACCCCTGAATGCTCTGGGTGAAGGTGTGAG	489
QY	472	GAGAGGCCCTTCAAGCCCCGAGGTGATCCCCATGTTCAACGGCCCTGAAGCGAGGGCGCCAC	531
Db	490	GAGAGGCCCTTCTCCCCCTGAGGTGATCCCCCATGTTCTCTGCCCTGTCTGAGGGGTCCACC	549
QY	532	CCCCGAGACCTGAACACGATGTTGAACACCGTGGCGGGCCACCAAGCCCGCATGCAATG	591
Db	550	CCCCGAGACCTGAACACCATGCTGAACACAGTGGGGGGCCATCAGGCTGCCATGCAATG	609
QY	592	CTGAAGGACACCATCAACGAGGAGCGCGGAGTGGGACCGCCTGCACCCCGTCAGGCC	651
Db	610	CTGAAGGAGACCATCAATGAGGAGCTGTGATGGGACAGGCTGCATCTGTGCACAGCT	669
QY	652	GGCCCGCTGGCCCCCGGCAGATGCGCGACCCCGCGCAGCGACATCGCCGGCGCCACC	711
Db	670	GGCCCGCATTGGCCCCCGGCAGATGAGGAGGCCAGGGGCTCTGACATTGCTGGCACCA	729
QY	712	AGCACCTTGACGAGACAGATCGCTGATGACCAACCCCGCTGCCCTGGGGCGAC	771
Db	730	TCCACCTTCAGGAGCAGATTGGCTGATGACCAACAACCCCGCATCCCTGTGGGGGAA	789
QY	772	ATCTACAAGCGGTGGATCATCTGGGCTGAACAAGATCGTGGATGTACACCCCGTG	831
Db	790	ATCTACAAGGTGGATCATCTGGGCTGAACAAGATGTGAGATGTACTCCCCCACC	849
QY	832	AGCATCTTGACATCCGCGCAGGGCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTC	891
Db	850	TCCATCTTGACATCAAGCAGGGCCCCAAGAGCCCTTCAAGGACTATGTGACAGGTTCC	909
QY	892	TTCAGAACCCTTGCGCGCCCGAGCAGGCCACCCAGGACGTGAAGAACTGGATGACGAGACC	951
Db	910	TACAGAACCCTGAGGGCTGAGCAGGCCCTCCAGGAGGTGAAGAACTGGATGACAGAGACC	969
QY	952	CTGCTGTGTCAGAAACGCCAACCCCGACTGCAAGAACCATCTTCGCGCTCTCGGCCCCGGC	1011
Db	970	CTGCTGTGTCAGAAATGCCAACCTTGACTGCAAGAACCATCTTGAAGGCCCTTGGGCCCTGCT	1029
QY	1012	GCCACCTTGAGAGATGATGACCGCTGCGCAGGGGCGTGGCGGGCCCCCGGCCACAAGGCC	1071
Db	1030	GCCACCTTGAGAGATGATGACCGCTGCGCAGGGGCGTGGGGGGCCCTGTGTACAAGGCC	1089
QY	1072	CGCGTGTGCGCGAGGGCGATGAGCCAGG---CCAACAGCGTGAACATCATGATGCAGAG	1128
Db	1090	AGGCTGTGCGTGAAGGCCATGTCCAGGTGACCAACTCCGCCACCATCATGATGCAGAGG	1149
QY	1129	AGCAACTTCAAGGGGCCCCCGCGCAACGTCAAGTGTTCAACTGCGGCAAGAGGGCCAC	1188
Db	1150	GGCAACTTCAAGAACAGAGAAAGACAGTGAAGTGTTCAACTGTGCAAGGTGGGCCAC	1209
QY	1189	ATCGCCAAGAACTGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCGGCAAGAGGGC	1248
Db	1210	ATTGCCAAGAACTGTAGGGCCCCCAGAGAAAGGGCTGTGGAAGTGTGCAAGAGGGC	1269
QY	1249	CACCATGTAAGAACTGCACCGAGCGCCAGGCCAACTTCTGTGGCAAGATCTGGCCAGC	1308
Db	1270	CACCATGTAAGAACTGCATGAAGGCAAGGCCAACTTCTGTGGCAAAATCTGGCCCTCC	1329
QY	1309	CACAAGGCGCGCCCCGGCAAATTCTGTGCAAAACGACGAGCCCGCGCGCCCCACCGTG	1368
Db	1330	CACAAGGCGCGCTGCGCAATTCTTCAGTCCAGGCTTAAGCCACAGCGCTTCCCGAG	1389

QY 1369 CCCACCGCCCCCGCCGAGACTTCGCTTCGAGGAGACCACCCCGCCCAAGCAG 1428  
 Db 1390 GAGTCCTC-----AGSTTGGGAGAGAAAGACCACCCCCAGAGAACAG 1437  
 QY 1429 GAGCCCAAGACCGCGAGCCCTACCGCGAGCCCTGCGCAGCCTGTTGGC 1488  
 Db 1438 GAGCCCATTTGACAGAAGAGCTGTAC-----CCCTGGCCTCCCTGAGGTCCCTGTTGGC 1491  
 QY 1489 AGCGGCCCCCTGAGCCAGTAA 1509  
 Db 1492 AACGACCCCTCTCTCCCAAGTAA 1512

```

RESULT 15
US-09-952-060-25
; Sequence 25, Application US/09952060
; Patent No. 673393
; GENERAL INFORMATION:
; APPLICANT: Emml, Emilio A.
; APPLICANT: Youll, Rima
; APPLICANT: Betc, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shaver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Castamiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 37474
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DNA encoding pMRkAd5 HIV-1 gag, coding
US-09-952-060-25

```

Query Match	Best Local Similarity	Matches 1210; Conservative	64.4%; Score 971.4; DB 4; Length 37474;	79.6%; Pred. No. 1.3e-138;	0; Mismatches 281; Indels 30; Gaps 4;
QY	1	ATGGCGCCCGCCGACGATCTGCGCGCGGAGAGCTGGACAGTGGAGATCCGC	60		
Db	1272	ATGGGTGCTAGGGCTTCTGTGCTGTCTGTGTGAGCTGGACAGTGGAGAGATCAGG	1331		
QY	61	CTGCGCCCCCGCGCAAGAGCACTACATGCTGAAGCACCTGTGTGGCCAGCCGAG	120		
Db	1332	CTGAGGCCTGTGGCAAGAGAGTACAAGCTAAAGCACATGTGTGGCTTCAGGAG	1391		
QY	121	CTGAGAGGCTTTCGCTCTGAACCCCGGCTGTGTGAGACCGCGGAGGCTGCAAGCATC	180		
Db	1392	CTGAGAGAGTTCGTGTGAACCCCTGGCTGTGTGAGACCTCTAGAGGGGTGACGGCAGATC	1451		
QY	181	ATGAAGCAGCTGCAGCCCGCCTTCAGACCGGACCGAGAGCTGCGCAGCTGTACAAC	240		
Db	1452	CTGGGGCAGCTCCAGCCCTCCCTGCAAAACAGGCTGTGAGAGCTGAGGTCCCTGTACAAC	1511		
QY	241	ACCGTGGCACCCCTGTACTGCGTGCACGCGCGCATCGAGTCCGCGACACCAAGAGGCC	300		

Db	1512	ACAGTGGCTACCCCTGTACTGTGTGCAACAGAGATTTGATGTGAAGGACACCAAGGAGGCC	1571
OY	301	CTGACCAAGATCGAGAGAGACAGAACAACTCCAGCAGAGAACCAGCAGGC-----	353
Db	1572	CTGAGAGAGATTGAGAGAGACAGAACAACTCCAAAGAGAGGCCCAAGCAGGCTGCTGCT	1631
OY	354	--CAAGAGGCGCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACTTCAGGGC	411
Db	1632	GGCACAGGCACTCCAGCCAGGTGTCCAGAACTACCCCATTTGTGCAAGAACTCCAGGGC	1691
OY	412	CAGATGTGCAACGAGGCCATCAGCCCCCGCACCCCTGAACGCTTGGGTGAAGTATCGAG	471
Db	1692	CAGATGTGCAACGAGGCCATCTCCCCCGGACCCCTGAATGCTTGGGTGAAGTGTGAG	1751
OY	472	GAGAAAGGCTTCAAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCCACC	531
Db	1752	GAGAAAGGCTTCTCCCTGAGGTGATCCCATGTTCTCTGCGCTGTCTGAGGGTGCACC	1811
OY	532	CCCCAGGACCTGAACACAGATGTTGAACACCGTGGGGCGGCCACAGGCCGCATGCAGATG	591
Db	1812	CCCCAGGACCTGAACACAGTCTGAACACAGTGGGGGGCCATCAGGCTGCATGCAGATG	1871
OY	592	CTGAAGGACACCATCAACGAGGAGGGCCGCGAGTGGGACCGCTTCACCCCGTGCAGGCC	651
Db	1872	CTGAAGGAGACCATCAATGAGGAGGCTGCTGAGTGGGACAGGCTGCATCTGTGCACGCT	1931
OY	652	GCCCCCGTGGCCCCCGGCGAGATGCGCGACCCCCCGGCGACGACATCGCCGGCGCCACC	711
Db	1932	GCCCCCATTTGCCCCCGGCGAGATGAGGGAGGCCCAAGGGCTCTGACATTTGCTGGCACCA	1991
OY	712	AGCACCTTCGAGGAGCAGATGCGCTTGATGACAGCAACCCCTCGTGCCGTGGCGAC	771
Db	1992	TCCACCTCCAGGAGCAGATTTGGCTGATGACCAACAAACCCCTCATCCTGTGGGGGAA	2051
OY	772	ATCTACAAGCGGTGATCATCTCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCCGTG	831
Db	2052	ATCTACAAGGTGATCATCTCTGGGCTTGAACAAGATTTGAGGATGTACTCCCCACC	2111
OY	832	AGCATCTTGAATATCCGCCAGGGGCCCAAGAGCCCTTCCGCACTACGTGACCGCTTC	891
Db	2112	TCCATCTTGACATCAGGCAAGGGCCCCAAGAGCCCTTCAGGGACTATGTGACAGGTTTC	2171
OY	892	TTCAAGACCTTCGCGCGCCGAGCAGGCCACCCAGGACGTGAAGAACTGGATGACCGAGACC	951
Db	2172	TACAAAGACCTTGAGGGCTGAGCAGGCTCCCAAGAGGTGAAGAACTGGATGACAGAGACC	2231
OY	952	CTGCTGTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCGGC	1011
Db	2232	CTGCTGTGCAAGATGCCAACCTCTAGTCAAGACCATCTGAAAGGCCCTCGGGCCCTGCT	2291
OY	1012	GCCACCTTGAGGAGATGATGACCGCTGCGCAGGGCGTGGGCGGCCCGGCCACAAGGCC	1071
Db	2292	GCCACCTTGAGGAGATGATGACAGCTGCGCAGGGGGTGGGGGGCCCTGCTCAAGGCC	2351
OY	1072	GCGGTGCTGGCGGAGGGCGATGAGCCAGG---CCAACAGCGTGAACATCATGATGCAGAG	1128
Db	2352	AGGGTGTGCTGAGGGCCATGTCCAGGTGACCAACTCCGCCACCATCATGATGCAGAGG	2411
OY	1129	AGCAACTTCAAGGGCCCCCGCGCAACGTCAAGTGTCTTCAACTGCGGCAAGGAGGGCCAC	1188
Db	2412	GCCAACTTCAGGAAACAGAGGAAAGACAGTGAAGTGTCTTCAACTGTGGCAAGGTGGCCAC	2471
OY	1189	ATCGCCAAAGAACTGCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCAGGCAAGGAGGC	1248
Db	2472	ATTGCCAAGAACTGTAGGGCCCCCAAGAGAGGGGCTGCTGGAAGTGTGGCAAGGAGGGC	2531
OY	1249	CACCAGATGAAGAACTGCACCGAGCGCCAGGCCAACTTCTTGGGGCAAGATCTGGCCACG	1308
Db	2532	CACCAGATGAAGAACTGCATGAGAGGCGAGGCCAACTTCTTGGGGCAAAATCTGGCCCTCC	2591
OY	1309	CACAAAGGCGCGCCCGCACTTCTCTGCAAGAACGCAAGCGACCCGCGCGCCCAACCGTG	1368
Db	2592	CACAAAGGCGCGCTTGCACTTCTCTCAAGTCCAGGCGCTGAGGCCACAGGCCCTCTCCGAG	2651

QY 1369 CCCACCGCCCCCGCCGAGAGCTTCGGCTTCGAGAGACCACCCCCGCCCAAGCAG 1428  
Db 2652 GAGTCCTTC-----AGGTTGGGGAGAGAGAAGACCACCCCCAGCAGAGCAG 2699  
QY 1429 GAGCCCAAGAGACCGCGAGCCCTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTGGC 1488  
Db 2700 GAGCCCATTTGACAAGAGAGCTGTAC-----CCCTGGCCCTCCCTGAGGTCCCTGTTGGC 2753  
QY 1489 AGCGGCCCCCTGAGCCAGTAA 1509  
Db 2754 AAGGACCCCTCTCCAGTAA 2774

Search completed: March 13, 2005, 00:22:49  
Job time : 273.651 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 16:34:11 ; Search time 5313.82 Seconds  
(without alignments)  
10809.355 Million cell updates/sec

Title: US-09-475-704A-4

Perfect score: 1509

Sequence: 1 atgggcgcgcgcgcgcagcat.....gcggcccccctgagccagtaa 1509

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*\n2: gb\_est2:\*\n3: gb\_hc:\*\n4: gb\_est3:\*\n5: gb\_est4:\*\n6: gb\_est5:\*\n7: gb\_est6:\*\n8: gb\_gss1:\*\n9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	87.4	5.8	1440	9 AG332984	AG332984 Mus muscu
2	84	5.6	1941	9 CL971508	CL971508 OsIFCC021
3	83.2	5.5	933	8 CC391778	CC391778 PUIJ707B
4	81.6	5.4	826	7 CN128942	CN128942 RHOH1_32
C 5	81.6	5.4	1219	6 CD500131	CD500131 CDA43-C08
6	81.6	5.4	1800	9 CL965300	CL965300 OsIFCC012
7	81.4	5.4	2031	9 CL974989	CL974989 OsIFCC042
8	81	5.4	749	5 BQ744212	BQ744212 WHEA113_A
9	80.6	5.3	1674	2 BE636712	BE636712 rockefe11
C 10	80.2	5.3	746	7 CK215205	CK215205 FGAS02715
11	79.8	5.3	2299	3 AY106831	AY106831 Zea mays
C 12	78.4	5.2	1126	7 CK211108	CK211108 FGAS02294
13	77.4	5.1	1064	7 CK163121	CK163121 FGAS01573
C 14	77	5.1	786	9 CG351928	CG351928 OGAG53TH
15	76.8	5.1	1138	7 CK163513	CK163513 FGAS01614
16	76.8	5.1	2151	9 CL972100	CL972100 OsIFCC041
17	76.4	5.1	759	2 BF259495	BF259495 HVSMEF001
C 18	76.4	5.1	869	7 CK159167	CK159167 FGAS04056
19	76.4	5.1	1308	3 AY104577	AY104577 Zea mays
20	76.4	5.1	1758	9 CL965380	CL965380 OsIFCC012
21	75.8	5.0	2757	9 CL982329	CL982329 OsIFSC047
22	75.4	5.0	883	9 CG274385	CG274385 OG2BH19TV
23	75	5.0	509	4 BG314500	BG314500 WHE2495_E
24	74.8	5.0	1047	9 CL967097	CL967097 OsIFCC014

C 25	74.4	4.9	1170	6 CD496623	CD496623 CDA23-C11
26	74.4	4.9	2598	3 AY103647	AY103647 Zea mays
27	74.2	4.9	649	6 CB870888	CB870888 HC15J04W
28	74	4.9	637	4 B1960118	B1960118 HVSMEH002
29	73.8	4.9	562	2 BF484304	BF484304 WHE2321_F
C 30	73.6	4.9	3249	9 CL945510	CL945510 OsIFSB004
C 31	73.2	4.9	636	6 CD932185	CD932185 GR45_117C
C 32	73.2	4.9	694	4 B261281	B261281 B261281
33	73	4.8	2629	9 CL955065	CL955065 OsJRU000
34	72.6	4.8	575	5 BQ842001	BQ842001 WHE2987_G
35	72.6	4.8	601	4 BJ290713	BJ290713 BJ290713
36	72.6	4.8	652	4 BJ292215	BJ292215 BJ292215
C 37	72.6	4.8	683	6 CD905815	CD905815 G468_102P
C 38	72.6	4.8	700	4 BJ295604	BJ295604 BJ295604
C 39	72.6	4.8	712	5 BQ842240	BQ842240 WHE2990_E
C 40	72.6	4.8	718	4 BJ297319	BJ297319 BJ297319
C 41	72.6	4.8	753	4 BJ298801	BJ298801 BJ298801
42	72.6	4.8	1150	7 CK161908	CK161908 FGAS01448
43	72.6	4.8	1197	7 CK161927	CK161927 FGAS01450
44	72.6	4.8	2256	9 CL946580	CL946580 OsIFSB005
45	72.6	4.8	2697	9 CL952258	CL952258 OsJRU000

#### ALIGNMENTS

RESULT 1  
AG332984/c  
LOCUS  
DEFINITION  
AG332984 1440 bp DNA linear GSS 02-JUN-2004  
Mus musculus molossinus DNA, clone:MSMG01-123F20.T7, genomic survey  
sequence.

ACCESSION  
AG332984  
VERSION  
AG332984.1 GI:47906294  
KEYWORDS  
GSS.  
SOURCE  
Mus musculus molossinus  
ORGANISM  
Mus musculus molossinus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the mouse BAC library MSMG01. For BAC  
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
Tsukuba Institute, Bio Resource Center,  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp  
PRIMERS  
Sequencing : T7  
LIBRARY : pBACE3.6  
Vector : EcoRI  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.

FEATURES  
source  
Location/Qualifiers

1..1440  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMG01-123F20.T7"  
/sex="male"  
/tissue\_type="mixture of kidney and spleen"  
/clone\_lib="MSMG01 Mouse Male BAC Library"

ORIGIN

Query Match 5.8%; Score 87.4; DB 9; Length 1440;  
Best Local Similarity 43.7%; Pred. No. 9e-06;  
Matches 424; Conservative 0; Mismatches 543; Indels 4; Gaps 2;

QY CGCCCTGAGGAGGGGCGCCACCCCGGAGACCTGAACACATGTTGAACACCGTGGCGG 569  
DB |||||  
1188 CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1129

QY 570 CCACGAGCGCGCGCATGACATGCTGAAGACACCATCAAGAGAGGCGCGCGAGTGGGA 629  
DB |||||  
1128 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1069

QY 630 CGGCTGCAACCGCGTGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 689  
DB |||||  
1068 CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1009

QY 690 CAGCGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 749  
DB |||||  
1008 CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 949

QY 750 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 809  
DB |||||  
948 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 890

QY 810 CGTGGCGATGTACAGCGCGCGTGAAGCATCTGACATCGCGCGCGCGCGCGCGCGCGCTT 869  
DB |||||  
889 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 830

QY 870 CCGCGACTACGTGAGACCGCTTCTCAAGACCTGCGCGCGCGAGAGCGCGCGCGCGCGCGT 929  
DB |||||  
829 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 770

QY 930 GAAGAACTGATGACCGAGACCCCTGCTGTGCAAGACCGCAACCCCGCATGCGAAGACCAT 989  
DB |||||  
769 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 710

QY 990 CCGCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1049  
DB |||||  
709 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 653

QY 1050 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1109  
DB |||||  
652 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 593

QY 1110 GAACATCATGATGCAAGAGCAACTTCAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1169  
DB |||||  
592 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 533

QY 1170 CTGCGGCAAGGAGGCGCATCGCGCAAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1229  
DB |||||  
532 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 473

QY 1230 GAAGTGGCGCAAGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1289  
DB |||||  
472 CAACACAGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 413

QY 1290 GGGCAAGATCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1349  
DB |||||  
412 AAAAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353

QY 1350 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1409  
DB |||||  
352 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 293

QY 1410 CACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1469  
DB |||||  
292 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233

QY 1470 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1480  
DB |||||  
232 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 222

RESULT 2  
CL971508  
LOCUS  
DEFINITION  
CL971508  
OS:Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.  
VERSION  
CL971508  
KEYWORDS  
SOURCE  
ORGANISM  
GSS.  
Oryza sativa (indica cultivar-group)  
Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 1941)  
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,  
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,  
Wong, G. K. S., Deng, X. W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and  
its comparison to Arabidopsis  
Unpublished (2004)  
CONTACT: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.  
Location/Qualifiers  
1. 1941  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_11b="Oryza sativa Expressed Library"  
/note="Oryza sativa exon trapped genomic sequences"

ORIGIN

Query Match 5.6%; Score 84; DB 9; Length 1941;  
Best Local Similarity 44.4%; Pred. No. 3.8e-05;  
Matches 584; Conservative 0; Mismatches 700; Indels 30; Gaps 5;

QY 98 ACCTGTGTGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 157  
DB |||||  
512 AGTGTATGCGCATCATCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571

QY 158 CCGCGGAGGCGTCAAGAGCATGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 217  
DB |||||  
572 AGCGAGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628

QY 218 AGGAGTGGCGAGCGCTGTACACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 277  
DB |||||  
629 TCGATGTGTGCTGTCTACCATGAGAGGCGCATCTTCAAGGTCAAGCGCGCGCGCGCG 688

QY 278 AGTCCGCGACCAAGAGGCGCTTGACAGATCGAGAGGAGAGAGAGAGAGAGAGAGAGTCC 337  
DB |||||  
689 AACACCACTGGCGCGCGAGGAGCTTGACACCGCATGGTGAACCACTTGTGAGAGAGT 748

QY 338 AGAAGACCGAGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 397  
DB |||||  
749 TCAAGCGCAAGAACAGAGGAGCATCAGCGCGCAACCGCGCGCGCGCGCGCGCGCGCG 808

QY 398 AGAAGTGAAGGCGCGAGATGTGACAGGCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCG 457  
DB |||||  
809 CGGCGTGAAGGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 868

QY 458 TGAAGTGAAGGAGAGAGGCGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514  
DB |||||  
869 ACTGCTGTACAGAGGCGCATGACTTCAACAGAGCATCAAGAGGCGCGCGCGTTCAGAGAG 928

QY 515 TGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 574

```
Db      929 TCACATGACCTGTTCCGCAAGTGCATGAGCCCGTGAGAAAGTGCTCCGCGACGCCA 988
Qy      575 AGCCCGCATGACAGATGTGAAGACACCATCAACGAGAGGCGCCGAGTGGAGCCGC 634
Db      989 AGATGACAAAGACAGCGGTGACGACGTGCTGTCGGCGGGCTCCACCCGCATCCCA 1048
Qy      635 TGACCCCGGTGACAGCGCGCGCGTGGCGCGCCGAGATGGCGGACCCCGCGCAGCG 694
Db      1049 AGGTGACAGACTGTCTCAAGACTTCTTCAACGCAAGAGAGCTTGCAAGACATCAACC 1108
Qy      695 ACATCGCGCGCGCCACACGACACCTCGACAGACAGATCGCCTGATGACGACCAACCCC 754
Db      1109 CCGACGAGCGCGGTGCGTACGCGCGCGCGCTCCAGCGCCCATCTGAGCGCGCAGGGCA 1168
Qy      755 CCGTCCCGTGGCGACACTTCAAGCGGTGGATCATCTGGGCTGAACAAGATCGTGC 814
Db      1169 ACGAGAAAGTGCAGGACCTGCTCTGCTGCAGCTCACCCG-----CTGTCCCTCG 1219
Qy      815 GGATGTACAGCCCCCGTGAACATCTGACATCCGCCAGGCGCCCAAGAGCCCTTCCCGC 874
Db      1220 GCCTGAGACCGCGCGCGCGGTATGACCGTGTCTATCCCGAAGAACACCATCCCGA 1279
Qy      875 ACTACGTGAGCCGCTTCTTCAAGACCTCGCGCGCGCGAGCAGCGCCACCGACGTGAAGA 934
Db      1280 CCAAGAAAGAGCAGGTCTTCTCCACCTACTCCGACAAACGACCGCGCTCTGATCCAG 1339
Qy      935 ACTGATGACCGAGACCTGCTGTGCAAGAACGCCAACCCGAGTGAAGACCATCTGC 994
Db      1340 TGTACGAGGGCGGAGCGCGCGCAGGACCAAGAACACACCTCTCGCAAGTTGAGCTCT 1399
Qy      995 GCGCTCTCGGCGCGCGCGCACCTGTGAGAGATGATGACCGCTGCCAGGCGGTGGCG 1054
Db      1400 CCGGCACTCCACCGCGCGCGCGCGCGCTGCCAGATCAAGTCTGC-----TTGC 1450
Qy      1055 GCGCGCGCACAAAGCGCGCGGTGTGCGCGAGGCGATGAGCCAGGCCAACAGCGTGAACA 1114
Db      1451 ACATCGACGCCAACCGGCACTCTGAACGTGTCCGCGAGACAAACGCGGGCAGAGA 1510
Qy      1115 TCATGATGCAAGAGCAACTTCAAGGCGCGCGCGCAAGTCAAGTCTTCAACTGCG 1174
Db      1511 ACAAGATCAACCATCAACCAAGCAAGGCGCGGTGAGCAAGAGAGATGAGAAAGTGG 1570
Qy      1175 GCAAGAGGCGCCACATCGCCCAAGAACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAGT 1234
Db      1571 TGCAAGAGCGGAGAAAGTCAAGGCGCGAGACGAGAGCAAAAG-----AAGGTGG 1624
Qy      1235 GCGGCAAGAGGCGCGCACAGATGAAGACTGCACCGAGCGCCCAACTTCTTGCGCA 1294
Db      1625 ACGCAAGAAAGCGCGCTCGAGAACTACGCTTCAACATGCGCAACACCATCAAGGACGAGA 1684
Qy      1295 AGATCTGGCGCGCGCACAGAGGCGCGCGCGCAACTTCTGCAAGAACCGCAGCGAGCGCG 1354
Db      1685 AGATCGCTCCAAAGCTCGCGCGCGCGCGACGACAAAGAGAGATCGAGGACGCTGATGGCG 1744
Qy      1355 CCGCGCGCACCGGTGCGCACCGCGCGCGCGCGCGAGAGCTTCCGCTTCAAGAGA 1408
Db      1745 CCATCAGCTGGCTCGACACCAACGAGCTCGCGAGCGCGAGGCGAGTTCAAGAGCA 1798

RESULT 3
LOCUS   CC391778          933 bp      DNA      linear      GSS 19-MAY-2003
DEFINITION   PUH1J707B_ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTA458K20,
              genomic survey sequence.
ACCESSION   CC391778
VERSION     CC391778.1  GI:30871868
KEYWORDS    GSS.
SOURCE      Zea mays
            Zea mays
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
```

```
REFERENCE 1 (bases 1 to 933)
AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
TITLE     Maize Genomics Consortium
JOURNAL   Unpublished (2003)
COMMENT   Other_GSSs: PUH1J707D
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.

FEATURES
Source     location/Qualifiers
            1..933
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBTA458K20"
            /clone_1ib="ZM_0.6_1.0_KB"
            /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
            cot selected genomic DNA library"

ORIGIN
Query Match      5.5%; Score 83.2; DB 8; Length 933;
Best Local Similarity 45.7%; Pred. No.5.1e-05;
Matches 329; Conservative 0; Mismatches 388; Indels 3; Gaps 1;

Qy      189 GCTGACGCCCGCCCTGACAGCCGCGCACCGAGAGCTGCGCAGCCTGTACACACCGTGGC 248
Db      17  GCCGAGGCGCGCGGTGTCCGGGCACCCGAGAAAGCGCGCTCGAAGCAAGCAGCGTGTGT 76
Qy      249 CACCTGTACTGCGTGCACCGCGGCATCGAGGTCCGCGCAACCAAGAGGCCCTTGACAA 308
Db      77  CAGGTGACACAGCACCCAGAGAGTTCGACGACGCGCTCAAGGCCGCCCAAGAACCGGCTGGT 136
Qy      309 GATCGAGAGAGGACAGAACAAAGTCCCAAGACAGAAAGCCCAAGAGGCCGCGACGG 368
Db      137 GGTGTGAGAGTTCGCGCGGAGCGACAGCAGAGAGACAGCAGATCTTACCCGACCATGT 196
Qy      369 CAAGGTGAGCCAGAACTAACCCCATCTGTGCAAGAACTGTGCAAGGCGCCAGATGTGCAACGAGC 428
Db      197 GCAGCTCAAGCCGACCTGCGCGGCGAGCTTGCACTTCTGTGCTGCTTCTGCGCGACGAGCTC 256
Qy      429 CATCAGCCCCCGCACCTGAAACGCTGGGTGAAGGTGATCGAGAGAAAGCCTTCAAGCC 488
Db      257 GGAGGCCACCGAGAGAGCTGTTCCGCGGAGAGGCGGTCAAGAAAGTGCCTCACTTCAACTT 316
Qy      489 CGAGGTGATCCCCCATGTTCACCGCGCTGAGCGAGGCGCCACCCCGCAGAGACT--GAA 545
Db      317 CTACAAAGGCGCGCGAGAAAGGTGCACGAGAGAGAGGCCATCGGCCGAGCGCTCGCGGG 376
Qy      546 CACGATGTGAACACCGTGGGCGCGCGCACAGGCGCGCATGCAAGATCTGAAGGACACCAT 605
Db      377 CGACGTCTCTACTACGCGCGACAGCACTCGCGGTGTGAGCTGCACTGCGGAGGA 436
Qy      606 CAACGAGAGGCGCGCGAGTGGAGCGCGCTGCACCCCGTGACGCGCGCGCGCGCGCGCG 665
Db      437 CGTGAGGCGCTCATTCAGACGACCGCGCGCGGACAAAGGCAAGCTGCTGTGCTGAGCT 496
Qy      666 CGGCGAGATGCGCGACCGCGCGCGCGCGAGCAGCATCGCGCGCGCGCACAGCAACCTGCAGGA 725
Db      497 GGGCTTCAAGCAGCTGCGCGCGCGCGCTGCAAGGTGTAACCCACCGTGTGAAGCTGTGCGG 556
Qy      726 GCAGATCGCTGTGATGACCAAGCAACCCCGCTGCGCGGTGGCGGACATCTCAAGCGGTG 785
Db      557 GTCCATGTGTGACAAACCGCTTTCGCGCGCATGAACGCGGACGAGAAAGACAGCTGAT 616
Qy      786 GATCATCTGGGCGCTGAACAAAGATGTGCGGATGTACAGCCCGTGAGCATCTTGACAT 845
```





OY 225 GCGCAGCCTGTACAAACACCGTGGCCACCTGTACTGCTGACGCGCGCATCGAGTCCG 284  
DB 327 GCTGGGGGTGTGACGAGGTAAAGAGCAGAGTGGAGCTGATCAGCGCGCAGCTGAAGAA 386  
OY 285 CGACACCAAGAGAGCCCTTGAACAAGATCGAGAGAGAGCAACAAGTCCACGAGAAGAC 344  
DB 387 GCGGAGCAAGAGCGGACGAGAGACGACAGACATGAGCTGGCCATGACCTGATGATGAT 446  
OY 345 CCAGCAGGCGCAAGAGAGCGGACGAGAGTGAAGCCAGAACTAACCTCGTGCAGAACT 404  
DB 447 CTTGACAGAGCAAGAGAGAGAGCGCA--CAACGCCGACCGCCCATCTGAGCCGCT 503  
OY 405 GCAGGGCCAGATGGTGCACCAAGCCATCAGCCCCCGACCTGAACGCTGGGTGAAGT 464  
DB 504 CGCCAGCGCCTCCAGCTCCAGAGCCTCGCCGACCTCCGCGCAGACCATGGCCATCAA 563  
OY 465 GATCGAGAGAGAGCCCTTACGCCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGG 524  
DB 564 GAACTCATCAACGACCAACGAGTCCGACTCCACCAACCAAGATTGTGACCTCTCCACCG 623  
OY 525 CGCCACCCCGCAGGACCTGAACAAGATGTTGAACACCGTGGCGGCGCACAGCGCCCAT 584  
DB 624 CTTCAAGGCGCATCGCGCGCTGACGAGAGAAACATCTCGGCGAGCTTTCATCCCA 683  
OY 585 GCAGATGCTGAAGAGACCACTCAACGAGAGGCGCGCAGAGTGGAGCCGCTGACCCGCT 644  
DB 684 GTACCTGAGAGAGTGCCCTTCTCTCATGATCCCAACGACTTCTCTGCCCCATCTCCCT 743  
OY 645 GCAGGCGCGCGCGCTGGCGCGCGCGCGCAGATGCGCGACCCCGCGCAGCGACATCGCGG 704  
DB 744 CGAGATCATGACCGATCCCACTTATGAACGCGCGCAGCATCCAGAAATGCTGACGCAAG 803  
OY 705 CGCCACCAAGACCCCTGACAGAGCAGATCGCTGATGACCAACCCCGCGTGGCGCGT 764  
DB 804 GCAGAGGAGCTGCGCCCAAGACGACGAGC--GCTGGGCGCATCTCTCTGCGCGCA 860  
OY 765 GGGCGACATCTCAAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGAGTACAG 824  
DB 861 CTACGCGCTCAAGAACCTCATCATGATGATGTCGACAAAGAGTGAAGATCCACTC 920  
OY 825 CCCCCTGAGCATCTTGAACATCCGCCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGA 884  
DB 921 CGGCGACCGCGCGCGCGGAGCCACCGGAGACCCCAAGGTGTCTATCCCAAGCTGTCAA 980  
OY 885 CCGCTTCTTCAAGACCTTGGCGCGCGCAGAGCGCCACCGAGAGCTGAAGAACTGATG-- 942  
DB 981 GGACCTTCTCTCGCCCAACTTGAAGTGAAGCGGAGGCGCTCAAGAGATCCGAGCGCT 1040  
OY 943 -----ACCGAGACCCCTGCTGTGTCAGAGAGCCCAACCCCGACTGCAAGACCAT 989  
DB 1041 GTCCAAGAGAGAACCCCGAGAGAGAGGCTGTCTGTACCGACCAACCGCGCATCCCGCGCT 1100  
OY 990 CTTGCGCGCTCTCGCGCGCGCGCGCGCCACCTTGAAGAGATGATGACCGCTGCGAGGGCGT 1049  
DB 1101 CATCGGCTTCTCTCTTACCCCGACAGAGAGATGACAGAGAACCGCTCACTGCTGCT 1160  
OY 1050 GGGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGAGGCGGATGAGCGCAAGCGCAAGCGT 1109  
DB 1161 CAACCTTCTCATCGAGAGGCGCAACAGCTCTCTCATGCGCAGGGGCGGGCCATCCCTT 1220  
OY 1110 GAACATCATGATGACAGAGAGAGCAACTTCAAG 1141  
DB 1221 GATCATCGAGCTGTCTAGAGAACGCGCAGCGTGG 1252

RESULT 7  
LOCUS CL974989 2031 bp DNA linear GSS 21-SEP-2004  
DEFINITION OsIFCC042895 Oryza sativa Express Library Oryza sativa (indica  
cultivar-group) genomic, genomic survey sequence.  
ACCESSION CL974989  
VERSION CL974989.1 GI:52404497  
KEYWORDS GSS.

SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
AUTHORS 1 (bases 1 to 2031)  
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,  
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,  
Wong, G. K. S., Deng, X. W. and Wang, J.  
TITLE An analysis of transcriptional regulation of the rice genome and  
its comparison to Arabidopsis  
JOURNAL Unpublished (2004)  
COMMENT Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.  
FEATURES  
Location/Qualifiers  
1..2031  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_1b="Oryza sativa Express Library"  
/note="Oryza sativa exon trapped genomic sequences"  
ORIGIN  
Query Match 5.4%; Score 81.4; DB 9; Length 2031;  
Best Local Similarity 44.0%; Pred. No. 0.00011;  
Matches 496; Conservative 0; Mismatches 616; Indels 15; Gaps 3;  
OY 313 GAGAGAGAGCAACAAGTCCCAAGAGAGACCCAGAGGCGCAAGAGCGGCAAG 372  
DB 280 GAGGCGCGCAAGAACAGAGCGGCGCGCAACCCGAGAGCATCTACAGCGCAAGCGG 339  
OY 373 GTGAGCAGAACTAACCCATGCTGACAGAACTGCAAGGCGCGAGATGTCACCAAGCCATC 432  
DB 340 CTGATCGCGCGCAGATTCTCGAGCGCGAGGTGACAGCGCATGAAGCTGCTGCGGTT 399  
OY 433 AGCCCCCGCACTTGAACGCTTGGGTGAAGTATCGAGAGAGAGGCTTCAAGCCCGAG 492  
DB 400 GCGTGTGTCAGACCGAAGCGGACCGGACGTGCGGTGAGTGAAGAGCGGCGAG-- 457  
OY 493 GTGATCCCAATGTTCAACGCGCTTGAAGCGAGGCGCAACCCCGAGACCTGAACACGATG 552  
DB 458 ----TGCGCGGTTCAGCC-----GAGAGAGGTGAGCGCATGTGTCAAGCGGATG 507  
OY 553 TTGAACACCGTGGGCGCGCAACAGGCGCGCATGACAGATGCTGAAGACCATCAAGAG 612  
DB 508 AAGAGAGCGCGAGGCTTACTCTGCGAGAGGTACACCGCGCGGTGACCGTCCG 567  
OY 613 GAGCGCGCGAGTGGAGCGCGCTGCAACCGGTGAGAGCGGCGCGCGCGCGCGCGAG 672  
DB 568 GCCTACTTCAACGACGCGCAGCGGACGCAACAGAGAGCGCGCGGTATGCGCGGCTC 627  
OY 673 ATGCGGACCGCGCGCGCGAGCATGCGCGCGCGCGCGCAAGACCTGACAGAGAGATC 732  
DB 628 ACCGTGACCGCATCATCAAGAGCGCGCGCGCGCGCGCGCATGCGCATGACAGAG 687  
OY 733 GCCTGATGACAGCAACCGCGCGCGCGCGCGCGCATCTCAAGCGGTGATGATC 792  
DB 688 AAGGCGCGAGAGAGAGCGTCTGTCTTGAACCTCGCGCGCGCGCGTTCAGAGTCAAGC 747  
OY 793 CTGGGCTGAACAAGATGCTGCGATGTAAGCCCGTGAAGCATCTGAGCATCCGCGAG 852  
DB 748 ATCTTGCCATGACAAAGGCGGTGTGAGAGTCTTTCACCAACGCGGACACCACTTC 807  
OY 853 GGGCCCAAGAGAGCCCTTCCGCGACTACGTGAGACCGCTTCTTCAAGACCTTGGCGCGAG 912  
DB 808 GCGGCGAGAGACTTGAACCAAGCGCTCATGAGCACTTGTCAAGGTATCCGCGGAGAG 867

OY		913	CAGGCCACCCGAGCAGTGAAGA	CTGGATGACCAGAACCCCTGTGTGTGACAGACGCCAAC	972	
Db		868	CACGGCCCGCATCA	CCGGCGCACGGCGCGCGCTGGGCMAAGCTCCGCCGAGTGCGAG	927	
OY		973	CCCAGCTGCAA	GACCATCTTCGCGCGCTCTCGGCCCCGGCGCCACCTTGAGAGATGATG	1032	
Db		928	CGCG--	-CCAAGCGCGGCTCAGCAA	CAGCACAGGTGCGCTGAGGTGAGTCCCTG	984
OY		1033	ACGGCTTC	CAGGGCGTGGGCGGCGCCGACAAAGGCCCGCGCTGTGGCCGAGCGCATG	1092	
Db		985	TTCGACGGCGT	CGACTTTGCGAGCGCTCTCCGGGGCGCGGTTCGAGGAGCTCAACAGC	1044	
OY		1093	AGCCAGGCCCA	CACAGCGTGAACATATGATGCAGAAAGCAATTCAAGGGCCCCCGCGC	1152	
Db		1045	GACCTCTTCA	AAGAAGATGTGTCCGGTGAAGGAAGGCCATGGCGGACGCCCGCTGAGC	1104	
OY		1153	AAGCTCAAGT	GCTTCAACTGCGGCAAGAGGGCCCATGCGCCAAAGAACTGCCCGCGCCCC	1212	
Db		1105	AAGGGCGACAT	CGAAGATCGTCTCGTGGCGGACGACCCAGATCCCCAAGSTGACG	1164	
OY		1213	CGCAAGAAGGG	CTGTGGAAGTGGCGCAAGAGGGCCACCAAGATGAAGACTGCAACGAG	1272	
Db		1165	CAGCTGCTCA	AGGACTACTTCGGTGGCAAGAGGCCAACCGCGCGTCAACCCCGACGAG	1224	
OY		1273	CGCCAGGCCAA	CTTCTGTGGCAAGATCTGGCCAGCCACAAGGGCGCGCCCGCAACTTC	1332	
Db		1225	GCGTGGCGTA	CGGCGCGCGGTGCAGGCAAGCATCATCAGCGGCCACGTGCAGAGAAC	1284	
OY		1333	CTGCAGAACCG	CAGCGAGCCCGCGCCCCCAACCGTGGCCACCGCCCCCGCGCGAGAGC	1392	
Db		1285	ACGAGAGCAT	GTATCTCTTCGACGTGCGCGCGCTCACCTCGGCTGAGAGCGGCCGC	1344	
OY		1393	TTCGCTTC	GAGAGAACCAACCCCGCCCCCAAGCAGAGAGCCCAAGGA	1439	
Db		1345	GCGGTGATGG	CCAAGCTGATCCGCGCAACACAGGTGTGCGCAGAA	1391	

RESULT 8	
BQ744212	
LOCUS	749 bp mRNA linear EST 17-JUL-2002
DEFINITION	WHE4113_A01 A01ZS wheat salt-stressed root cDNA library Triticum aestivum CDNA clone WHE4113_A01_A01, mRNA sequence.
ACCESSION	BQ744212
VERSION	BQ744212.1 GI:21890999
KEYWORDS	EST.
SOURCE	Triticum aestivum (bread wheat)
ORGANISM	Triticum aestivum
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum. 1 (bases 1 to 749)
AUTHORS	Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J., Iazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
TITLE	The structure and function of the expressed portion of the wheat genomes - Salt-stressed root cDNA library
JOURNAL	Unpublished (2002)
COMMENT	Contact: Olin Anderson US Department of Agriculture, Agricultural Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: oanderson@pw.usda.gov Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: SK primer.

```
FEATURES
source
Location/Qualifiers
1..749
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
```

```

/clone="WHE4113_A01_A01"
/tissue_type="Roots"
/dev_stage="Full tillering"
/lab_host="E. coli SOLR"
/clone_1ib="Wheat salt-stressed root cDNA library"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid
pbluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Hydropionic
plants grown to full tillering stage were treated with 150
mM NaCl for either 12 hours or 7 days. Root tissues of the
plants subjected to both types of treatment were collected
separately at University of California, Davis (E. Akhunov
and K. Deal in J. Dvorak's lab). Total RNA was prepared
separately from the two samples (12h and 7day treatments),
and equal amount of RNA was then pooled. PolyA RNA was
purified from the pooled RNA, a cDNA library was made, and
the cDNA clones were in vivo excised to give pbluescript
SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak)
at the University of California, Davis. Colony plating,
plasmid DNA preparations and DNA sequencing were performed
in the OD Anderson lab (all other authors)."

```

```

ORIGIN
Query Match      5.4%; Score 81; DB 5; Length 749;
Best Local Similarity 48.4%; Pred. No. 0.00013;
Matches 293; Conservative 0; Mismatches 300; Indels 12; Gaps 2;

```

QY		269	CCGGCATCGAGTCCGCGACACCAGAAGAGGCCCTTGACAAGATCAGAGGAGCAGACA	328
Db		8	CCTTCAGCCTCGTCGACGCCAAGCCGCCGCCCAAGCACCCGCCTTCGGGCCCAAGTGGC	67
QY		329	AGTCCAGCAGAAAGACCCAGCAGGCCAAGGAGGCCGACGGCAAGGTGAAGCCAGAACTAAC	388
Db		68	GCTTCAACCAAGCGGCCCCAGCTCCCCAGCGCCGCGACGAGAGGTGAGGCCAAGCGCC	127
QY		389	CCATCGTGCAAACTTGACGGGCCAGATGTGTGCAACAGGCCATCAGCCCCCGCACCTGA	448
Db		128	GCGAGGCCGAGAAAGAGCGCGCCCGCCGAGCGCCACTAACAGAACCAACGCTCCAC	187
QY		449	ACGCGTTGGG--TGAAGTGATCGAGGAGAAAGGCGCTTCAGCCCCGAGGTGATCCCATGT	505
Db		188	ACCACCAGGGCTTCGCGGCAACAGACAGAACATCGCCAAAGCCCTTCGTGACATCCAGC	247
QY		506	TACCGCGCCCTGAGCGAGGGCGGCCACCCCCCAGACCTGAACAAGATGTTGAACACCGTGG	565
Db		248	CCGACTGACCATCTCTGAGCAGATCCCCTTCGCCAACTTACACCAAGCTCTCCTTCGCGG	307
QY		566	GCGGCGCACCAAGGCGCCCATGTCAGATGCTGAAGGACACCATCAACGAGGAGCGCCGAGT	625
Db		308	TGCGCCGACCAAGCCCGAAGAGCTGTGCTGTGCGGCGCGTGCAGCTCGTAACGCCGCT	367
QY		626	GGGACCGCGCTGCACCCCGTGACAGGCCGCCCCCGGTGGCCCCCGGCCAGATGCGCGACCCC	685
Db		368	ACGACCGCGTCAACCCCAAGACGCGCCCGCGCTTCGAGCGCTTCAAAGAACCGCCAGTTCT	427
QY		686	GCGGCAGCGCATCGCCGCGGCCCAAGCACCTGC-----AGGACAGATCGCCT	736
Db		428	TCAAGATCAACCAACCGACGACCCCATCATCGCGCGCCTCGCCGAGAGGACAAAGGCCA	487
QY		737	GGA TGACCAAGACCCCCCGGTGCCCGGTGGGCGACATCTCAAGCGGTGATCATCTGG	796
Db		488	CGGCTTCGCCACCGACGCCATCTCGCCGCGCTCATGTGACGCCCCGACGATCTCT	547
QY		797	GCCTGAACAAGATCGTGCGGATGTACAGCCCGGTGAGCATCTTGACATCCGCGACGGCC	856
Db		548	CCTGGGACATGTGCTGTCAGCGCGTCCGCAACAAGCTTCTTGACAAGCGCGAGGGCT	607
QY		857	CCAAG 861	
Db		608	CCCAG 612	

RESULT 9  
BE636712

LOCUS BE636712 1674 bp mRNA linear EST 03-JAN-2002  
DEFINITION rockefeller.0.42 Mastigamoeba balamuthi lambda ZAP II library  
Mastigamoeba balamuthi cDNA similar to vacuolar ATP synthase  
subunit B (EC 3.6.1.3), mRNA sequence.  
BE636712  
ACCESSION BE636712 GI:9919823  
VERSION  
KEYWORDS  
SOURCE Mastigamoeba balamuthi  
ORGANISM Mastigamoeba balamuthi  
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
REFERENCE 1 (bases 1 to 1674)  
Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,  
Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and  
Philippe, H.  
The analysis of 100 genes supports the grouping of three highly  
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
21819461  
JOURNAL MEDLINE  
PUBMED 11830664  
COMMENT Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockefeller.edu  
Insert Length: 1674 Std Error: 0.00  
POLYA=Yes  
FEATURES  
source 1. 1674  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II library"  
/note="syn: Phreatamoeba balamuthi"  
ORIGIN  
Query Match 5.3%; Score 80.6; DB 2; Length 1674;  
Best Local Similarity 44.1%; Pred. No. 0.00015;  
Matches 615; Conservative 0; Mismatches 756; Indels 22; Gaps 6;  
QY 59 GCCTGCGCCCCGCGGCAAGACATGCTGAAGCACTGTGTGGCCAGCCGCG 118  
Db 136 GCGCTGTGTGCTGCTCTGACAAACATCAAGAAAGCCGCAATTGCAAGATCGTCAACTGA 195  
QY 119 AGCTGAGAGGGCTTCCGCTGAAACCCGCGCTGTGAGACCGCCGAGGGCTGCAAGAGA 178  
Db 196 CGCTGGGGGACGGCAGCGTGGCGCGGGCAGGTGCTGAGAGTGGCGGAAACAAGGCGC 255  
QY 179 TCATGAAGCAGCTGCAGCGCCGCTGAGACCGGCACGAGAGAGCTGCGCAGCTGTACA 238  
Db 256 TCGTGCAGGTTCTTGAGGGCACGTGGGCAATGACGCCAAGAACACGCACTGCGAGTTCA 315  
QY 239 ACACCGTGGCCACCCCTGTACTGCGTGAACGCGCGCATGAGGTCCGCGACACCAAGAA-- 296  
Db 316 CGGGCAACATCATGCGCATCCCGGTGACGAGGACATGCTGGGCGCGCTTCAACGGCT 375  
QY 297 -GGCCCTGACAAAGATGAGAGGAGAGCAAAACAAGTCCAGCAGAAACCCAGAGGCCA 355  
Db 376 CGGGCGCGCGCGTCCGCAAGGGCGCGCGGTGATGCGCGAGACTTCTCGACATCCAGG 435  
QY 356 AGGAGCGCGACGGCAAGGTGAGCCGAGAACTACCCCATGCTGACAGAACTGAGGCGCAGA 415  
Db 436 GCCAGCCCATCAACCCGTACAAAGCGTGTACCCCGAGAGAGATGATCCAGACGGGCATCA 495  
QY 416 TGGTGCAACGAGCCATCAAGCCCGGACCCCTGAAGCGCTGGTGAAGGTGATCAGAGAGA 475  
Db 496 GCGGCATCGACACGATGAAGTCTGATGCGCGGTGCGCAGAAAGATCCGCTGTTACGCGCG 555  
QY 476 AGGCTTTCAGCCCGGAGGTGATCCCATGTTTCAACGCGCCCTGAGGAGGGCGCCACCCCC 535  
Db 556 CCGGTCTGCGCGCAACGAGATTGCGCGCAGATCTGACAGGCGAGCGCGCTGTGCGCC 615  
QY 536 AGGACTGAACGATGTTGAACACCGGTGGCGGCCACGAGCGCGCATGACAGATGCTGA 595

Db 616 TCCAGAGCAGCAAGAGGTGACAGAGACACGAGAGCACTTCGCCATCGTGTTCGCGG 675  
QY 596 AGGACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGTGACAGCCGCGCC 655  
Db 676 CCATGGGTGTCAACATGAGAGACTGCAGGTTCTTCGCCCAGAGACTTCAGAGAGAACGGCT 735  
QY 656 CCGTGCGCGCGCGCGAGATGCGCGACCCCGCGGAGCGACATGCGCGCGCCACCAAGCA 715  
Db 736 CGATGAGCGGTGTACAGCTGTCTTCAACCTCGCCAACACCCGACGATCGAGCGCATCA 795  
QY 716 CCCTGACGAGCAGATGCGCCTGATGACCAAGCAACCCCGCGTGGCGGCGCATCT 775  
Db 796 TCACGCGCGCTGTG---GCGCTCACGACGCGCGAGTAACTCGCGTACAGTGCAGATGC 852  
QY 776 ACAAGCGGTGATCATCTGCGCT-----GAACAAGATCGTGGGATGTACAGCCCGC 829  
Db 853 ACGTGTGTGTCATCTGACGAGCATGTCTGTGTAACGCGATGCGCTGTGAGTCTCGG 912  
QY 830 TGAACATCTGACATCCGCGCAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCT 889  
Db 913 CGGCGCTGAGAGGTGCGCTGGCGGTGTTAACCCCGGTACATGTACAGCATCTCG 972  
QY 890 TCTTCAAGACCTGCGCGCGCGAGCAGCCACCCAGAGAGTGAAGAACTGATGACCGAGA 949  
Db 973 CGAACATCTACGAGCGTGCAGGCGGTG---TCAGGGCAGAGAACGGCTCGATCACGACG 1029  
QY 950 CCTGTGTGTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTTCGCGCCCG 1009  
Db 1030 TCCCATCTTCACGATGCCCAAGACGACATCAACCAACCCCATCTGACCTCACGGGCT 1089  
QY 1010 GCGCCACCTGAGAGAGATGATGACCGCTGCGAGGCGGTGGGCGCGCCGCAAGG 1069  
Db 1090 ACATCACCGAGGCGCAGATCTACGTGACAGGCACTGACAAACAGCAGATCTACCCGC 1149  
QY 1070 C---CGCGTGTGCGCGCGAGCGATGAGCCAGGCCAACAGCGTGAACATGATGACAGA 1126  
Db 1150 CATCAACGTGTGCGCTGCTGACCGGTGATGAAAGCGCCATCGCGGAGGCGATGA 1209  
QY 1127 AGAGCAATTCAAGGGCGCGCGCGCAACGTCAAGTCTTCAACTGCGCGCAAGAGGCGC 1186  
Db 1210 CGCGGACGACCACTCGAGGTGTGAAACCAAGCTGTACGCCAATACGCGCATCGCAAG 1269  
QY 1187 ACATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGCTGTGAAAGTGCAGCAAGAGG 1246  
Db 1270 ACAGATGGCCATGAAGGCTGTGCTGTGAGAGAGGCGCTGTGCGCGAGACAAGCTCG 1329  
QY 1247 GCCACGATGAAGAGACTGCAACGAGCGCGCAAGCTTCTGGGCAAGATGTGCGCA 1306  
Db 1330 CCTGAGTTCTCTGAGAACTTGAAGAGTTATCAAGCAGAGGCTTCAACGAGTGC 1389  
QY 1307 GCCAAGAGGCGCGCGCGCAACTTCTGTGAGAAACGCGAGAGCCCGCGCCACCGC 1366  
Db 1390 GCGACATCTAACAGTGTGCTGACATGCGCTGAGAGCTTGTGCG---CAAGTTCGCCAAG 1445  
QY 1367 TGCCCAACGCGCGCGCGCGAGAGCTTCCGCTTGAAGAGAACACCCCGCGCCCAAGC 1426  
Db 1446 AGCTCTTCAAGCGGTATCACGGAAGAACCTTCAACAAGTACTACAGCAGGAGCGCAAGN 1505  
QY 1427 AGGAGCCCAAGGA 1439  
Db 1506 AGGAGCACCAAGTA 1518  
RESULT 10  
CK215205/c 746 bp mRNA linear EST 09-DEC-2003  
LOCUS CK215205/1  
DEFINITION FGAS027158 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
aestivum cDNA, mRNA sequence.  
ACCESSION CK215205  
VERSION CK215205.1 GI:39621309  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 746)  
REFERENCE Allard, P., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
Penniket, C., Roach, J.L. and Sarhan, F.  
TITLE Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
JOURNAL Unpublished (2003)  
COMMENT Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_est@cs.usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [13,662].  
Plate: L6B016 row: K column: 15.  
Location/Qualifiers  
1..746  
FEATURES  
source

/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown  
(50%) and leaf (50%) tissues from wheat cultivar Norstar  
after short exposure times to low temperature in the light  
and in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20C from wheat cultivar Norstar after  
short exposure times to low temperature in the light and  
in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20, wheat plants were transferred to 4C  
in the light. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. The last 6 populations: After 7 days  
of growth at 20C, wheat plants were transferred to 4C in  
the dark. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. First strand synthesis in this  
library was done in the presence of methylated dCTP  
thereby protecting from internal cleavage with NciI. In  
addition, this library used a primer for second strand  
synthesis that annealed to an artificial sequence (RNA  
oligo) added before first strand synthesis. Therefore when  
sequences from EST generated from this library will be  
masked for vector and adaptor sequences, an additional  
masking step will have to be included to mask this RNA  
oligo that is common to all clones (sequence  
CGACTGAGCAGAGGACATGACATGACTGAAGAGTAGAA)."

ORIGIN  
Query Match 5.3%; Score 80.2; DB 7; Length 746;  
Best Local Similarity 49.7%; Pred. No. 0.00018;  
Matches 261; Conservative 0; Mismatches 258; Indels 6; Gaps 2;  
QY 191 TGCAGCCCCCTGCAGACCGGACCGAGAGCTGCCGACCTGTACAACACCGTGCGCA 250  
DB 670 TGCTGCTCAAGCTCATCGGCGCCAGAGAACCATGAGATCGCGCTTACACCGGGTACT 611  
QY 251 CCCTGTACTGCGTGACCGCGGCATCGAGGTCCGGACACCAAGAGGCGCTTGACACAGA 310  
DB 610 CCCTCTCTGCCACCGCGCTCGCATCCCGACGACGCGCACCATCTTGGCCATGACATCA 551  
QY 311 TCGAGAGGAGGACGAGCAAGTCCACGAGAGAACCCAGGCGCAAGGAGGCGGACGCGCA 370

DB 550 ACCGGAGAACTAGAGCTGGGGCTGCCGTGCATCGAGAGAGCGCGGTGGCAGACAGA 491  
QY 371 AGGTGAGCCAGAACTACCCCATCGTGCAGAGAACCTGACAGGGCCAGATGTTGACACAGGCCA 430  
DB 490 TCGACTTCGGCGAGGGCCCGCGCTGCCGTGCTGAGACGCGCTGCTGAGAGCAGAGGCCA 431  
QY 431 TCAGCCCCCGCACCTGAAACGCTGGGTGAAGTGAATCGAGAGAGAGGCTTCAGCCCCG 490  
DB 430 ACCACGGACACTTTCGACTTCTGTCTGTGAGACGCGCAGACAGACAACTACTCACTACC 371  
QY 491 AGTGATCCCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCGACGACTGAACAGA 550  
DB 370 ACGAGCGCTCATGAGCTCGTCAAGCTCGGCGCTCTCGGCTACGAC--AAACGCG 314  
QY 551 TGTGAACACCGTGGCGCGCCACGACCGCCCATGAGATGCTGAAGACACCATCAAG 610  
DB 313 TCTGAACGGCTCCGTCGTGCTCCCGCGCGCCCGCCCATGCGCAAGTACATCCGCTACT 254  
QY 611 AGGAGCGCGCGAGTGGGACCGGCTGCACCCCGTGCAGGCGCGGCC--CGTGCCCCCG 667  
DB 253 ACCGCACTTCGTCCTCTGACCTCAACAGAGGCCCTCGCGCCGACGCGCGCTGAGATCT 194  
QY 668 GCCAGATGCGGACCGCCCGCGGACGACATCGCCGCGCCACCA 712  
DB 193 GCCAGCTCCCGTCGCGGACGCGCATCCTCTGCGCGCGGCCA 149

RESULT 11  
AY106831  
LOCUS 2299 bp mRNA linear HTC 16-OCT-2002  
DEFINITION Zea mays PC0142079 mRNA sequence.  
ACCESSION AY106831  
VERSION AY106831.1 GI:21209909  
KEYWORDS HTC.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 2299)  
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,  
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.  
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of  
Overgo Probes  
JOURNAL Unpublished (2002)  
REFERENCE 2 (bases 1 to 2299)  
Coe, E.H.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of  
Missouri, Columbia, MO 65211, USA  
COMMENT If you are interested in getting corresponding physical clones,  
these are publicly available from ZmDB and may be found by BLAST  
searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,  
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the  
maize cDNA sequences is either Virginia Walbot, Stanford or Pat  
Schnable, Iowa State, then clones may be requested from ZmDB:  
www.zmdb.iastate.edu.

FEATURES  
source  
1..2299  
Location/Qualifiers  
1..2299  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="MaizeDB:638374"  
/db\_xref="taxon:4577"  
/clone\_lib="Maize Mapping Project/DuPont Consensus  
Library"  
/note="this sequence is part of a project of EST  
assemblies resulting from the application of public  
contigs to seed DuPont contigs; this resource was  
assembled by DuPont as part of a collaboration for the  
overgo addressing of BACs in conjunction with the Maize  
Mapping Project"

ORIGIN

Query Match		5.3%;	Score 79.8;	DB 3;	Length 2299;
Best Local Similarity		43.9%;	Pred. No. 0.00022;		
Matches 495;		Conservative 0;	Mismatches 617;	Indels 15;	Gaps 3;
QY	122	TTGAGGGCTTGGCCCTGAAACCCCGGCTGTGTGAGACCGCGGAGGGCTGCAAGCATCA	181		
DB	767	TCGAGGCGCGGCATCATGAGACATCTGGATGCGAGCTCTTATGAGCAGCGCA	826		
QY	182	TGAAGCAGTCGACACCCGCTGACAGCGGCAACGAGAGCTGCGAGCTGTACACACA	241		
DB	827	AGGAGGTGAAGCCCATGAGACCGCTGTGTAAGCCGAAGACAGAGTACGCGCTCCGCA	886		
QY	242	CCGTGGCCACCTGTACTGCTGACAGCGCGGCATGAGGTCCGCGACACCAAGAGGCC	301		
DB	887	CGTCGCCGAGTGGCTCGGCCCCAGATCGAGGTCTCCGCGCCGCGCACCAAGTCCATCG	946		
QY	302	TGACAAAGATCGAGAGAGACAGAACAGTCCAGCAGAGAACCCAGAGGCCAAGAGAG	361		
DB	947	AGCAGAGGTCAACTCGGTCAACAGAACCTGTGATGACGTCCACCGTGGCAAGCGC	1006		
QY	362	CCGACGGCAAGGTGAGCCAGAACTACCCCATGTCAGAACTTGACAGGCGCATGTGTC	421		
DB	1007	TGCACGGCGCAACTTCCAGGGCACCCCGATCGCGTGTTCATGACAAAGCCCGCTCG	1066		
QY	422	ACCAGGCCATCAGCCCCCGCACCTGAAAGCCTGGGTGAGGTGATCGAGAGAGAGCCT	481		
DB	1067	CCGTGCGCAACATCGGCAAGCTCATGTGCGCCAGTTCTCCAGTGTGTCACAGATTCT	1126		
QY	482	TCAGCCCCGAGGTGATCCCATGTTTACCGCCCTGACGAGCGAGCGCCACCCCGCAGACC	541		
DB	1127	ACAACAAGGGGCTGACCTTCAACCTGCGCGGACGCGCAACCGAGCTGAGACTACGGGT	1186		
QY	542	TGAACACGATGTTGAACACCGTGGGCGGCGCACAGCGCCCATGACAGATGCTGAAGACA	601		
DB	1187	TCAAGGGCAACGAGATCGCATGGCGTCTTACTGCTCCGAGCTCCAGTACTTGCCCACT	1246		
QY	602	CCATCA-----ACGAGAGGCGCGCGAGTGGACCGCTGACCCCGCTGACAGCGCGCC	655		
DB	1247	CCATCAACAACGATGTCAGAGCGCGGAGACACAAACAGACGTGAATCTCCCTCGGCC	1306		
QY	656	CCGTGGCCCCCGGCGCAGATGCGGACCGCGCGGACGACATCGCGCGCGCCACGACA	715		
DB	1307	TCGTCTCGCCAGAAAGACCGCGAGGCGGTGACATCTTCAAGCTATGTGTCAAGT	1366		
QY	716	CCCTGACGAGCAGATGCTCTGATGACCAACACCCCGCTGCGCGTGGGCGACATCT	775		
DB	1367	ACATGTTCGCGCTGTGCGCAGGCGCTGACCTGCGCACTCGAGAGAACTCAAGAGCG	1426		
QY	776	ACAAGCGGTGATCATCTCTGGGCTGAAACAAGATCTGCGGATGTACAGCCCGTAGACA	835		
DB	1427	CCGTCAAGAGCTGCTGATGCGGCTGCGGAGAGGTGCTGACCAACCTCGCGCGCG	1486		
QY	836	TCCTGACATCCGCGCAGGCGCCCAAGAGCGCTTCCGCGACTACGTGACCGCTTCTCA	895		
DB	1487	ACCTCAACAGCGCGCGCTTACGAGAAAGCCCTGCTTCAACCGCATTCGACCGCGAGCGG	1546		
QY	896	AGACCTGCGCGCGCAGAGCGCCACCAAGACGTGAAGAACTGATGACCGAGA-----	949		
DB	1547	TGTACGGGTACTACGACGACCCCTGACGCGCAACTGCGCCCTGTGTAAGAAATCCGGG	1606		
QY	950	CCCTGTGTGTGACAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCCCCG	1009		
DB	1607	CCGTGTGTGTGACACGCGCTTACGAGCGGCGGAGCGCGAGAGACGCGCGGTCCG	1666		
QY	1010	GCGCCACCTTGAGAGATGATGACCGCTGCGAGGCGTGGCGCGCCCGGCCACAAGG	1069		
DB	1667	TGTTCTTCAAGATCAACAGTTTCGAGAGAGAGCTGCGGAGCGCTGCCAGGAGATGG	1726		
QY	1070	CCCGGTGTGTGCGCGAGCGATGAGCCAGGCCAACAGCGTGAACATCATGATGACAGAGA	1129		
DB	1727	AGGCGCGCGCGCGTGTGAGACCGCGCGCGCGCGCATCGCCAACAGATCAAGAGA	1786		

QY	1130	GCAACTCAAGGCGCCCGCGCAACGTCAAGTGTCTCACTGCGGCAAGAGGCCACA	1189		
DB	1787	GCCGCTGTA---CCGCTGTACCGCTTATCCGCCAGAGACCTCGCGCGCTGTACTGA	1843		
QY	1190	TCGCAAGAACTGCGCGCCCCCGCAAGAGGCTGTGAGTGC	1236		
DB	1844	CCGCGCAAGAGCTCAAGTCCCCCGCGAGAGGTGCAACAGGTTC	1890		

RESULT 12  
CK211108/c  
LOCUS  
DEFINITION  
FGAS022942 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
aestivum cDNA, mRNA sequence.  
ACCESSION  
CK211108  
VERSION  
CK211108.1 GI:39617217  
KEYWORDS  
SOURCE  
ORGANISM  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1126)

REFERENCE  
AUTHORS  
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Linke, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
Penniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_est@cs.usask.ca

TITLE  
JOURNAL  
COMMENT  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [12,750].  
Plate: L6B001 row: A column: 20.

FEATURES  
source  
1..1126  
Location/Qualifiers

/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown  
(50%) and leaf (50%) tissues from wheat cultivar Norstar  
after short exposure times to low temperature in the light  
and in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20°C from wheat cultivar Norstar after  
short exposure times to low temperature in the light and  
in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20, wheat plants were transferred to 4°C  
in the light. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. The last 6 populations: After 7 days  
of growth at 20°C, wheat plants were transferred to 4°C in  
the dark. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. First strand synthesis in this  
library was done in the presence of methylated dCTP  
thereby protecting from internal cleavage with NotI. In  
addition, this library used a primer for second strand  
synthesis that annealed to an artificial sequence (RNA  
oligo) added before first strand synthesis. Therefore when  
sequences from EST generated from this library will be  
masked for vector and adaptor sequences, an additional

masking step will have to be included to mask this RNA  
oligo that is common to all clones (sequence  
CGACTGAGCAGCAGGACACTGACATGACTGAGAGAGTAGAAA)."

# ORIGIN

Query Match 5.2%; Score 78.4; DB 7; Length 1126;  
Best Local Similarity 48.1%; Pred. No. 0.00038;  
Matches 285; Conservative 0; Mismatches 301; Indels 6; Gaps 2;

QY 124 GAGGCTTGGCCCTGAACCCCGGCTGTGAGACCGCGAGGCTGCAAGCATATG 183  
Db 763 GAGATCAACCGCAACCAACCATGAAACCTGATGACGACGTCGGCGAGAGGCGCAATTG 704  
QY 184 AAGCAGCTGACAGCCCGCCCTGACAGACCGGCAACGAGAGCTGCGCAGCTTACAACACC 243  
Db 703 CTCAACATGCTGTCTCAGCTCATCGCGCCAGAAGACCATGAGATCGCGCTTACACC 644  
QY 244 GTGGCCACCTGTACTGCGGTGACGCGGCATCGAGTCCGCGACACCAAGAGGCGCTG 303  
Db 643 GGCTACTCCCTCTCCGCGACCGGCTCGGCATCCCGGACGAGCGACCATCTTGCGCATG 584  
QY 304 GACAGATCGAGAGAGACAGAACAGTCCAGCAGAGACCCAGCAGGCCAAGAGGCC 363  
Db 583 GACATCAACCGGAGAACTACGAGCTGGGGCTGCGGTGATCGAGAGGCGCGGTGGCG 524  
QY 364 GACGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAAGACTGCAAGGCGCATGTGAC 423  
Db 523 CACAAGATCGACTTCCGCGAGGGCGCGGCGCTGCGGTGATGAGCGCGCTGTGAGAGAC 464  
QY 424 CAGGCCATCAGCCCCCGACCCCTGAAACGCTGGGTGAAGGTGATCGAGAGAGAGGCTTC 483  
Db 463 GAGGCCAACCAACGCACTTTCGACTTCTTCTGAGCGCCGACAGAGAACAACTACTTC 404  
QY 484 AGCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGGAGGCGCCACCCCGCAGACTG 543  
Db 403 AACTACCAACGAGCGCTCATGAAAGCTCGTCAAGCTCGGCGGCTCTCTGCGCTACGAC 347  
QY 544 AACACGATGTTGAACACCGTGGGCGGCAACGAGCGCCATGCAAGTGTGAAGACACC 603  
Db 346 AACACGCTCTGGAACGGCTCCGTCGTGCTCCCGCGAGCGCCCATGCGCAAGTACATC 287  
QY 604 ATCAACGAGGAGGCGCGGAGTGGAGCGGCTGCAACCGGCTGCAAGCGGCGCC 660  
Db 286 CGCTCTTACCGGACTTCTGCTCTGCACTCAACAGGCTCTGCGCGGAGCAGCGGCTC 227  
QY 661 GCGCCGCGGCGAGATGCGGAGACCCCGCGGCAAGCAGATGCGCGGCGGCAACCA 712  
Db 226 GAGATCTGCCAGCTCCCGCTCGGCGAGCGGATCAACCTCTGCGCGCGGCGCA 175

RESULT 13  
LOCUS CK163121 1064 bp mRNA linear EST 05-DEC-2003

DEFINITION FGAS015739 Triticum aestivum FGAS: Library 4 Gate 8 Triticum  
aestivum cDNA, mRNA sequence.

ACCESSION CK163121  
VERSION CK163121.1 GI:38993028  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

REFERENCE  
AUTHORS Allard, F., Crosby, W. L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L. D., Laroche, A.,  
Link, M. G., McCarthy, E. L., Monroy, A., Muzak, I., Nilsson, D.,  
Penniket, C., Roach, J. L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)

TITLE JOURNAL  
COMMENT Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science

1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_est@cs.usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [1,834].  
Plate: L4B07 row: C column: 11.  
Location/Qualifiers  
1..1064

## FEATURES

source

/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_1b="Triticum aestivum FGAS: Library 4 Gate 8"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6;  
Conditions for growth: Seeds were germinated in a  
water-saturated mix (50% black earth and 50% Promix) in a  
growth chamber for 7 days under an irradiance of 200 mmol  
m-2 sec-1. The temperature was maintained at 20 degrees C  
with a 15-hr photoperiod under a relative humidity of 70%.  
After this period watering of plants was stopped. Four  
time points were sampled during a two week period; the  
first after wilting was observed and the last, two weeks  
later, consisted of live crown and leaf tissue (leaf  
tissue that was yellow was not included in sampled  
material). First strand synthesis in this library was done  
in the presence of methylated dCTP thereby protecting from  
internal cleavage with NotI."

## ORIGIN

Query Match 5.1%; Score 77.4; DB 7; Length 1064;  
Best Local Similarity 48.0%; Pred. No. 0.00058;  
Matches 284; Conservative 0; Mismatches 302; Indels 6; Gaps 2;

QY 124 GAGGCTTGGCCCTGAACCCCGGCTGTGAGACCGCGAGGCTGCAAGCATATG 183  
Db 282 GAGATCAACCGCAACCAACCATGAAACCTGATGACGACGTCGGCGGAGAGGCGCAATTG 341  
QY 184 AAGCAGCTGACAGCCCGCCCTGACAGACCGGCAACGAGAGCTGCGCAGCTGTACAACACC 243  
Db 342 CTCAACATGCTGTCTCAAGCTCATGCGCGCCAGAGAACCATGAGATCGCGGTACACC 401  
QY 244 GTGGCCACCTGTACTGCGGTGACGCGGCATCGAGTCCGCGACACCAAGAGGCGCTG 303  
Db 402 GGCTACTCCCTCTCCGCGACCGGCTCGGCATCCCGGACGAGCGGACCATCTTGCGCATG 461  
QY 304 GACAGATCGAGAGAGCAGAACAGTCCAGCAGAGAACCCAGCAGGCCAAGAGAGGCC 363  
Db 462 GACATCAACCGGAGAACTACGAGCTGGGGCTGCGGTGATGAGAGAGCGCGGTGGCG 521  
QY 364 GACGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAAGACTGCAAGGCGCGCATGTGAC 423  
Db 522 CACAAGATCGACTTCCGCGAGGGCGCGGCTGCGGTGATGAGCGGCTGTGAGAGAC 581  
QY 424 CAGGCCATCAGCCCCCGACCCCTGAAACGCTGGGTGAAGGTGATGAGAGAGAGGCTTC 483  
Db 582 GAGGCCAACCAACGCACTTTCGACTTCTTCTGAGCGCGGAGCGGAGAGAGAGAGAGAG 641  
QY 484 AGCCCGAGGTGATCCCATGTTTCAACCGGCTGAGAGAGGCGGCGGCGGCGGAGACTG 543  
Db 642 AACTACCAACGAGCGCTCATGAAAGCTGTCMAAGCTTGGCGGCTCTCTGCGCTACGAC 698  
QY 544 AACAGATGTTGAACACCGTGGGCGGCAACGAGCGCCATGCAAGTGTGAAGACACC 603  
Db 699 AACAGCTCTGGAACGGCTCCGTCGTGCTCCCGCGAGCGCCCATGCGCAAGTACATC 758  
QY 604 ATCAACGAGGAGCGCGGAGTGGAGCGGCTGCAACCGGCTGAGCGGCGGCGGCGGCGG 660  
Db 759 CGCTACTACCGGACTTCTGCTGCACTCAACAGAGGCGGCTGCGGCGGAGCAGCGGCTC 818



m-2 sec-1. The temperature was maintained at 20 degrees C with a 15-hr photoperiod under a relative humidity of 70%. After this period watering of plants was stopped. Four time points were sampled during a two week period; the first after wilting was observed and the last, two weeks later, consisted of live crown and leaf tissue (leaf tissue that was yellow was not included in sampled material). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

ORIGIN

Query Match 5.1%; Score 76.8; DB 7; Length 1138;  
Best Local Similarity 47.9%; Pred. No. 0.00074;  
Matches 283; Conservative 0; Mismatches 303; Indels 5; Gaps 2;

```

QY 124 GAGGCTTGGCCCTGAACCCCGCTGTGAGACCGCGAGGCTGCAAGCATCATG 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 GAGATCAACCGCAACCAACCATGAACTGATGACGACGTGCGGAGAGGCCAGTTC 291

QY 184 AAGCAGTCAGCCCGCTGACAGCGGACCGAGAGCTGCGACGCTGTACAACACC 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 CTCAACATGCTGCTCAAGCTCATCGGCGCAAGAAGACCATGAGATCGCGCTTACACC 351

QY 244 GTGGCCACCTGTACTGCTGTCACGCGGACATGAGTCCGACACCAAGAGGCCCTG 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 GGCTACTCCCTGCTCGCACCGCGCTGCGCATCCCGACGACGACCATCTTGCCATG 411

QY 304 GACAAGATCGAGAGAGACAGAACTCCAGCAGAAAGCCAGACGCCAAGAGGCC 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 GACATCAACCGGAGAACTACGAGCTGGGGCTGCCGTGCATCGAAGAGCCGCGTGGCG 471

QY 364 GACGGCAAGTGAGCCAGAACTACCCCATGCTGACAGAACTGCAGGGCCAGATGTCAC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 CACAAGATCGACTTCGCGGAGGGCCCGCGCTGCGCCGTGAGCGCGTGTGAGGAC 531

QY 424 CAGGCCATCAGCCCCCGCACCTGTAAGCTTGGTGAAGTGATCGAAGAGAGCCTTC 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 GAGGCCAACCAACGGAACCTTCACTTCTTCTGTGACGCCGACAGAACTACCTC 591

QY 484 AGCCCCGAGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCAGACCTG 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 AACTACCAACGAGCGCTCATGAGCTGTCAAGTCCGCGGCTCTCTGCGCTACGAC--- 648

QY 544 AACACGATGTTGAACACCGTGGCGGACCAAGCGCCCATGCAGATGCTGAAGACACC 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AACACCTCTGGAACGCTCCGTCTGTCTGCTGAGCTCAACAGGCTCTGCGCGCAATAC 708

QY 604 ATCAACGAGAGGCGCGCGAGTGGACCGCTGCAACCCCTGCAAGCGCGG--CCCGTGG 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 CGCTACTACCGGACTTCTGCTGAGCTCAACAGGCTCTGCGCGNCAACAGCGCGTGG 768

QY 662 CCCCCGCGCAGATGCGGACCCCGCGGCAAGCAATGCGCGGCGCACCA 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AGATCTGCCAGCTCCCGTGGCGAGCGCATCACCTCTGCGCGCGCGCCA 819

```

Search completed: March 13, 2005, 00:13:49  
Job time : 5318.82 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:04:55 ; Search time 52.7746 Seconds  
(without alignments)  
5502.310 Million cell updates/sec

Title: US-09-475-704A-4  
Perfect score: 2840  
Sequence: 1 atcgggcgccgcgcgcagcat.....gcggccctgagccagctaa 1509

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool\_p/US09475704/runat.10032005.140222.14839/app\_query.fasta.1.3342  
-DB=PIR\_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704 @CGN 1.1.160 @runat.10032005.140222.14839 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2252.5	79.3	497	1	FOLJND	gag polyprotein -
2	2247	79.1	502	1	FOVMA2	gag polyprotein -
3	2237	78.8	512	1	FOVWH3	gag polyprotein -
4	2233	78.6	500	1	FOVWH4	gag polyprotein -
5	2232.5	78.6	501	2	SS4377	gag polyprotein -
6	2230	78.5	500	1	FOVWL	gag polyprotein -
7	2227	78.4	500	1	A44001	gag polyprotein -
8	2227	78.4	500	2	S33979	gag polyprotein -
9	2204.5	77.6	505	2	T01667	gag polyprotein -
10	2198	77.4	498	2	T09436	gag polyprotein -
11	2179.5	76.7	506	1	A38068	gag polyprotein -
12	2018.5	71.1	478	1	FOVWL	gag polyprotein -
13	1963	69.1	508	1	FOLJSI	gag polyprotein -
14	1403.5	49.4	554	2	S46346	gag polyprotein -

15	1396	49.2	510	1	FOLJTM	gag polyprotein -
16	1396	49.2	521	2	S08435	gag polyprotein -
17	1390	48.9	521	1	FOLJCA	gag polyprotein -
18	1388	48.9	521	2	S53091	gag polyprotein -
19	1385	48.8	521	1	FOLJST	gag polyprotein -
20	1381.5	48.6	507	2	S04237	gag polyprotein -
21	1376	48.5	507	2	T11559	gag polyprotein -
22	1374.5	48.4	522	1	FOLJG2	gag polyprotein -
23	1372	48.3	521	2	S12152	gag polyprotein -
24	1369.5	48.2	522	1	FOLJG3	gag polyprotein -
25	1367.5	48.2	506	1	FOLJG5	gag polyprotein -
26	1353.5	47.7	519	1	FOLJG4	gag polyprotein -
27	1344	47.3	519	1	FOLJG4	gag polyprotein -
28	1228.5	43.3	502	2	S28080	gag polyprotein -
29	713.5	25.1	171	2	S52929	GAG protein - huma
30	700	24.6	146	2	S60708	GAG protein - huma
31	698	24.6	146	2	S60702	GAG protein - huma
32	693	24.4	146	2	S60703	GAG protein - huma
33	691	24.3	146	2	S60698	GAG protein - huma
34	691	24.3	146	2	S60704	GAG protein - huma
35	689	24.3	146	2	S60699	GAG protein - huma
36	688	24.2	146	2	S60697	GAG protein - huma
37	678	23.9	146	2	S60700	GAG protein - huma
38	576.5	20.3	212	2	S03070	GAG protein - huma
39	468	16.5	449	2	A45557	matrix, capsid, nu
40	467	16.4	450	1	FOLJFP	GAG polyprotein -
41	467	16.4	450	2	S23819	GAG protein - fe11
42	460	16.2	486	1	FOLJFV	GAG polyprotein -
43	453	16.0	450	2	S25162	GAG protein - fe11
44	429.5	15.1	1106	2	JQ0405	hypothetical 119.5
45	420	14.8	85	2	S49086	GAG protein - huma

ALIGNMENTS

RESULT 1  
FOLJND  
gag polyprotein - human immunodeficiency virus type 1 (isolate NDK)  
N/Alternate names: core polyprotein  
N/Contains: core protein p15; core protein p17; core protein p24  
C/Species: human immunodeficiency virus type 1, HIV-1  
A/Note: host Homo sapiens (man)  
C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C/Accession: JQ0065  
R/Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.;  
Gene 81, 275-284, 1989  
A/Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immuno  
A/Reference number: JQ0065; MUID:90034200; PMID:2806917  
A/Accession: JQ0065  
A/Molecule type: DNA  
A/Residues: 1-497 <SPI>  
A/Cross-references: UNIPROT:P18800; GB:M27323; NID:G328154; PIDN:AAA44868.1; PID:G328157  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polyprotein  
C/Keywords: AIDS; core protein; immunodeficiency; polyprotein  
F/1-129/Product: core protein p17 #status predicted <C17>  
F/130-389/Product: core protein p24 #status predicted <C24>  
F/390-497/Product: core protein p15 #status predicted <C15>

Alignment Scores:

Pred. No.: 1.82e-106 Length: 497  
Score: 2252.50 Matches: 423  
Percent Similarity: 91.49% Conservative: 39  
Best Local Similarity: 83.76% Mismatches: 32  
Query Match: 79.31% Indels: 11  
DB: 1 Gaps: 4

US-09-475-704A-4 (1-1509) x FOLJND (1-497)

QY 1 ATGGGGCGCCGCGCAGCATCTCGCGCGGAGAGCTGACAGTGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgIaSerValLeuSerGlyGlyIysLeuAspTrpGluArgIleArg 20

```
QY 61 CTGCGCCCCGCGCAAGACACTACATGCTGAAGCACTGTGTGGGCCAGCCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysTyrAlaLeuLysHisLeuIleTrrPalaSerArgGlu 40
QY 121 CTGAGAGGCTTCGCGCTGAACCCCGGCTGTGAGACCGCGCGAGGCTGCAAGCAGATC 180
Db 41 LeuGluArgPheThrLeuAsnProGlyLeuLeuGluTrrSerGluGlyCysLysGlnIle 60
QY 181 ATGAAGCACTGACAGCCCGCTGACAGCCGACCGAGAGCTGCGAGCTGTACAC 240
Db 61 IleGlyGlnLeuGlnProSerIleGlnTrrGlySerGluGluIleArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCTGTCAGACGCGGACATCGAGTCCGCGACACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGluArgIleGluValLysAspThrLysGluAla 100
QY 301 CTGAACAGATCGAGAGGAGCAGAAAGTCCAGACAGAAAGCCAGAGCCCAAGAG 360
Db 101 ValGluLysMetGluGluGlnGlnAsnLysSerLysLysThrGlnGlnAlaAlaAla 120
QY 361 GCCGACGCGCAAGGTGAGCGCAAGACTACCCCATCGTGCAAGCCTGAGGCGCAGATGGTG 420
Db 121 AspSerSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetVal 140
QY 421 CACCAAGCCATCAGCCCCCGCACCCCTGAAGCCTGGGTGAAGTGAATCGAGAGAGGCC 480
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrrValLysValIleGluGluLysAla 160
QY 481 TTCAGCCCCGAGGTGATATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCGCAGAC 540
Db 161 PheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY 541 CTGAACACGATGTTGAACACCGTGGGCGGCCACCAAGCCGCGCATGAGATGCTGAAGAC 600
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysGlu 200
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGGACCGCTGACCCCGTGAAGCGCGCCCGCTG 660
Db 201 ThrIleAsnAspGluAlaGluTrrAspArgLysHisProValHisAlaGlyProVal 220
QY 661 GCCCGCGCGCAGATGCGCGCGACCCCGCGCGCAGCAGCATCGCGCGCCACCAAGCCCTG 720
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240
QY 721 CAGAGCAGATCGCTGTGATGACCAACACCCCGCTGCGCGCGCAGCATCTTCAACAAG 780
Db 241 GlnGluGlnIleAlaTrrMetThrSerAsnProIleProValGlyGluIleTrrLys 260
QY 781 CGGTGGATCATCTGGGCTTGAACAGATCGTGGATGTACAGCCCGTGAGCATCTG 840
Db 261 ArgTrrPheIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIleLeu 280
QY 841 GACATCCGCGCAGGCGCCCAAGAGAGCGCTTCCGCGACTACGTGAGCGCTTCTTCAAGACC 900
Db 281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPheTrrLysThr 300
QY 901 CTGCGCGCGCGAGCAGGCCACCCAGACGTGAAGAACTGATGACCGAGACCTGCTGTG 960
Db 301 LeuArgAlaGluGlnAlaSerGlnAspValLysAsnTrrMetThrGluThrLeuLeuVal 320
QY 961 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCGCGCGCACCCCTG 1020
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGlnAlaThrLeu 340
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGCTGGGCGCGCCGCGCAAGGCGCGTGCTG 1080
Db 341 GlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360
QY 1081 GCCGAGGCGATGAGCCAG-----GCCAACAGCGGTGAACATCATGATGACAGAGCAAC 1134
Db 361 AlaGluAlaMetSerGlnValTrrGlySerAlaThrAlaValMetMetGlnArgGlyAsn 380
```

```
QY 1135 TTCAAGGGCCCCGCGCGCAACGTCAGTGTCTTCACTGCGGCAAGAGGCGCCATCGCC 1194
Db 381 PheLysGlyProArgLysSerIleLysCysPheAsnCysGlyLysGluGlyHisThrAla 400
QY 1195 AAGAACTGCGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGCAAGAGGCGCCACG 1254
Db 401 LysAsnCysArgAlaProArgLysLysGlyCysTrrLysCysGlyArgGluGlyHisGln 420
QY 1255 ATGAAGCACTGACCGAGCGCGCAAGCTTCTGCGCAAGTCTGGCCAGCCACCAAG 1314
Db 421 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrrProSerHisLys 440
QY 1315 GCGCGCGCGCGCAACTTCTGCGCAAGCCGACGCGAGCCCGCGCGCCACCGTGCACAC 1374
Db 441 GlyArgProGlyAsnPheLeuGlnSerArgProGlu-----ProThr 454
QY 1375 GCGCGCGCGCGCGAGAGCTTCCGCTTC---GAGAGACCAACCCCGCGCGCAAGAG 1431
Db 455 AlaProProAlaGluSerPheGlyPheGlyGluGluIleThrProSerGlnLysGlnGlu 474
QY 1432 CCGAAGACCGCGAGCGCTTACCGCGAGCGCCCTGACCGCGCTGCGCAGCCTGTCGAGAC 1491
Db 475 GlnLysAspLysGluLeuTyr-----ProLeuAlaSerLeuLysSerLeuPheGlyAsn 492
QY 1492 GCGCGCGCTGAGCCAG 1506
Db 493 AspProSerSerGln 497
```

## RESULT 2

FOVMA2  
gag polyprotein - human immunodeficiency virus type 1 (isolate ARV-2)  
N/Alternate names: core polyprotein  
N/Contains: core protein p15; core protein p17; core protein p24  
C/Species: human immunodeficiency virus type 1, HIV-1  
A/Note: host Homo sapiens (man)  
C/Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004  
C/Accession: A03947  
R/Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh  
Science 227, 484-492, 1985  
A/Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).  
A/Reference number: A04003; MUID:85090453; PMID:2578227  
A/Accession: A03947  
A/Molecule type: DNA  
A/Residues: 1-502 <SAN>  
A/Cross-references: UNIPROT:P03349; GB:K02007; NID:G328658; PIDN:AAB59875.1; PID:G328661  
C/Comment: Cleavage sites that yield the mature core proteins remain to be determined.  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polyprotein  
C/Keywords: AIDS; core protein; immunodeficiency; polyprotein  
F/1-134/Product: core protein p17 #status predicted <P17>  
F/135-393/Product: core protein p24 #status predicted <P24>  
F/394-502/Product: core protein p15 #status predicted <P15>

## Alignment Scores:

Pred. No.:	3.44e-106	Length:	502
Score:	2247.00	Matches:	424
Percent Similarity:	90.78%	Conservative:	39
Best local Similarity:	83.14%	Mismatches:	31
Query Match:	79.12%	Indels:	16
DB:	1	Gaps:	5

US-09-475-704A-4 (1-1509) x FOVMA2 (1-502)

```
QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGAGAGCTGACAACTGGAGAGATCCGC 60
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrrPglLysIleArg 20
QY 61 CTGCGCGCGCGCGCAAGACACTTACATGCTGAAGCACTGTGTGGGCCAGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysTyrLysLysLeuLysHisIleValTrrPalaSerArgGlu 40
QY 121 CTGAGGCTTCGCGCTTGAACCCCGGCTGTGAGACCGCGCGCTGCAAGCAGATC 180
```

[illegible]

```

Db      401 GlyHisIleAlaIysAsnCySArGAlAProArGlySLySGlyCyStrPrArCySGlyArg 420
QY      1243 GAGGGCCACCAATGAMGAGCTGCACCCGAGGCGCCAGGCCAACTTCCTGGGCAAGATCTGG 1302
Db      421 GluGlyHisGlnMetLysAspCyStrGluArGlnAlaAsnPhLeuGlyLysIleTrp 440
QY      1303 CCCAGCCACAAAGGGCCGCGCCCGGCAACTTCCTGCAGAACCGCAGCGCCGCGCGCCCC 1362
Db      441 ProSerTyrLysGlyArgProGlyAsnPhLeuGlnSerArGProGlu----- 456
QY      1363 ACCGTGCCCAACGCGCCCCCGCGCGAGAGCTTCGGCTTC-----GAGGAGACCAACCCCC 1416
Db      457 -----ProThrAlaProProGluGlnSerPheArgPheGlyGluGluLysThrThrPro 474
QY      1417 GCCCCCAAGCAGGAGGCCCAAGSACCGGAGCCCTACCGCGAGCCCTGACCGCCCTGCGC 1476
Db      475 SerGlnLysGlnGluProIleAspLysGlnLeuTyr-----ProLeuThrSerLeuArg 492
QY      1477 AGCCTGTTGGGCAGCGGCGCCCTTGAGCCAG 1506
Db      493 SerLeuPheGlyAsnAspProSerSerGln 502

```

### RESULT 3

gag polyprotein - human immunodeficiency virus type 1 (isolate HTLV-III, BH10)  
N;Alternate names: core polyprotein  
N;Contains: core protein p15; core protein p17; core protein p24  
C;Species: human immunodeficiency virus type 1, HIV-1  
A;Note: host Homo sapiens (man)  
C;Date: 17-May-1985 #sequence 17-May-1985 #text\_change 09-Jul-2004  
C;Accession: A03945  
R;Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K.J.; Starcich, B.; Josephs, S.F.; Dora  
nberger, J.A.; Papas, T.S.; Ghayeb, J.; Chang, N.T.; Gallo, R.C.; Wong-Staal, F.  
Nature 313, 277-284, 1985  
A;Title: Complete nucleotide sequence of the AIDS virus, HTLV-III.

A/Reference number: A93353; MUID:85111123; PMID:2578615  
 A/Accession: A03945  
 A/Molecule type: DNA  
 A/Residues: 1-512 <RAT>  
 A/Cross-references: UNIPROT:P03347; GB:M5654; GB:K02008; GB:K02009; GB:K02010; NID:g326  
 C/Comment: Cleavage sites that yield the mature core proteins remain to be determined.  
 C/Genetics:  
 A/Gene: gag  
 C/Superfamily: AIDS-related virus gag polyprotein  
 C/Keywords: AIDS; core protein; immunodeficiency; polyprotein  
 F/1-132/Product: core protein p17 #status predicted <P17>  
 F/133-391/Product: core protein p24 #status predicted <P24>  
 F/392-512/Product: core protein p15 #status predicted <P15>

### Alignment Scores:

Pred. No.:	1,1e-105	Length:	51
Score:	2237.00	Matches:	42
Percent Similarity:	89.69%	Conservative:	39
Best Local Similarity:	82.10%	Mismatches:	39
Query Match:	78.77%	Indels:	14
DB:	1	Gaps:	5

US-09-475-704A-4 (1-1509) x F0VWH3 (1-512)

[illegible]

Db 61 LeuGIleuGIInProSerLeuGIInThrGlySerGIuLeuAArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCCTGTACTGCGTGCAAGCGCGCATCGAGGTCCGCAACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGIuIleLysAspThrLysGIuAla 100  
QY 301 CTGACCAAGATCGAGAGGAGCAGAAACAAGTCCAGCAGAAAGACCAGAGGCCAAGAG 360  
Db 101 LeuAspLysIleGIuGIuGIuGIuAsnLysSerLysLysValAlaGlnGlnAlaAla 120  
QY 361 GCCGACGGC-----AAGGTAGCCGCAACTACCCCATCGTGCAAGCCTGACAGGC 411  
Db 121 AspThrGIyHisSerSerGIInValSerGIInAsnTyrProIleValGIInAsnIleGIu 140  
QY 412 CAGATGGTGCAACGAGGCCATCAGCCCGCGCAACCTGAACGCTGGGTGAAGTGAAG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleValLysValValGIu 160  
QY 472 GAGAAAGCCTTCAGCCCGCAGGTGATCCCATGTTCACCGCCCTGACGAGCGCGCAC 531  
Db 161 GluLysAlaPheSerProGIuValIleProMetPheSerAlaLeuSerGIuGIuAlaThr 180  
QY 532 CCCGAGACCTGAACACGATGTTGAACAACCGTGGCGCGCACAGCGCGCATGACATG 591  
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGIyGIyHisGlnAlaAlaMetGlnMet 200  
QY 592 CTGAAGACACCATCAACGAGAGGCGCGCGAGTGGACCGCCTGACCGCCGTGACGAG 651  
Db 201 LeuLysGIuThrIleAsnGIuGIuAlaIleGIuTyrAspArgValHisProValHisAla 220  
QY 652 GGGCCCGTGGCCCCCGCGCAGATGCGCGAAGCGCGCGCAGCAGACATGCGCGCGCAC 711  
Db 221 GIyProIleAlaProGIyGIuMetArgGIuProArgGIySerAspIleAlaGIyThrThr 240  
QY 712 AGCACCTTCAGAGAGACAGATCGCCTGATGACCAAGCAACCCCGCGTGGCGCGAC 771  
Db 241 SerThrLeuGIuGIuGIuIleGIyTyrMetThrAsnAsnProProIleProValGIyGIu 260  
QY 772 ATCTACAAAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGATGTACAGCCCGTG 831  
Db 261 IleTyrLysArgTyrIleIleLeuGIyLeuAsnLysIleValArgMetTyrSerProThr 280  
QY 832 AGCATCTTGACATCGCGCCAGAGGCGCGCAAGAGAGCCTTCGCGCATCAGTGACCGCTTC 891  
Db 281 SerIleLeuAspIleArgGIuGIyProLysGIuProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAAGACCTTGCGCGCGCGCAGAGCGCAACCCAGACGTGAAGAACTGATGACCGAGACC 951  
Db 301 TyrLysThrLeuArgAlaGIuGIuAlaSerGIuGIuValLysAsnThrMetThrGIuThr 320  
QY 952 CTGCTGTGTGCAAGAGCCCAACCCCGCACTGCAAGACCATCTGCGCGCTCTGGCGCGGC 1011  
Db 321 LeuLeuValGIInAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGIyProAla 340  
QY 1012 GCCACCTTGAGAGATGATGACCGCGCTGCCAGGCGGTGGCGCGCGCGCCCAAGGCC 1071  
Db 341 AlaThrLeuGIuGIuMetMetThrAlaCysGIuGIyValGIyGIyProGIyHisLysAla 360  
QY 1072 CCGGTGCTGGCCGAGCGCATGAGCCAG--GCCAACAAGCGTGAAACATCATGATGCAAG 1128  
Db 361 ArgValLeuAlaGIuAlaMetSerGIInValThrAsnThrAlaThrIleMetMetGIInArg 380  
QY 1129 AGCAACTTCAAGGCGCGCGCGCAACGTCAAGTGTCTCAACTGCGGCAAGAGGGCCAC 1188  
Db 381 GlyAsnPheArgAsnGIInArgLysMetValLysCysPheAsnCysGIyLysGIuGIyHis 400  
QY 1189 ATGCGCAAGACTGCGCGCGCGCGCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGGGGC 1248  
Db 401 ThrAlaArgAsnCysArgAlaProArgLysGIyCysTyrLysCysGIyLysGIuGIy 420  
QY 1249 CACCAAGATGAAGACTGCAACCGAGCGCGCAAGGCCAATTCTGGGCAAGATCGCGCCAGC 1308  
Db 421 HisGlnMetLysAspCysThrGIuArgGIInAlaAsnPheLeuGIyLysIleTyrProSer 440

QY 1309 CACAAGGGCGCGCGCGCAACTTCTGTGCAAGACCGCAGAGCGCGCGCGCCACCGTG 1368  
Db 441 TyrLysGIyArgProGIyAsnPheLeuGIInSerArgProGIuProThrAlaProPhe 460  
QY 1369 -----CCACCGCGCGCGCGCGCGAGAGCTTCCGC-----TTCGAG 1404  
Db 461 LeuGIInSerArgProGIuProThrAlaProProGIuGIuSerPheArgSerGIyValGIu 480  
QY 1405 GAGACCAACCGCGCGCGCGCGCAAGCAGAGCCCAAGAGCCGAGCCCTACCGGAGCCCTG 1464  
Db 481 ThrThrThrProProGIInLysGIuGIuProIleAspLysGIuLeuTyr-----ProLeu 498  
QY 1465 ACCGCGCTGCGAGCCTGTTCCGACGCGCGCGCGCTGAGCCAG 1506  
Db 499 ThrSerLeuArgSerLeuPheGIyAsnAspProSerSerGIIn 512

RESULT 4  
FOVWH4  
gag polyprotein - human immunodeficiency virus type 1 (isolate CDC-451)  
N;Alternate names: core polyprotein  
N;Contains: core protein p15; core protein p17; core protein p24  
C;Species: human immunodeficiency virus type 1, HIV-1  
C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 09-Jul-2004  
C;Accession: A25523  
R;Desai, S.M.; Kalyanaraman, V.S.; Casey, J.M.; Srinivasan, A.; Andersen, P.R.; Devare, Proc. Natl. Acad. Sci. U.S.A. 83, 8380-8384, 1986  
A;Title: Molecular cloning and primary nucleotide sequence analysis of a distinct human A;Reference number: A94136, MUID:87041461, PMID:3490666  
A;Accession: A25523  
A;Molecule type: DNA  
A;Residues: 1-500 <DES>  
A;Cross-references: UNIPROT:P05887; GB:M13136; NID:g326459; PIDN:AAA44306.1; PID:g326462  
C;Genetics:  
A;Gene: gag  
C;Superfamily: AIDS-related virus gag polyprotein  
C;Keywords: core protein; polyprotein  
F;1-132/Product: core protein p17 #status predicted <P17>  
F;133-391/Product: core protein p24 #status predicted <P24>  
F;392-500/Product: core protein p15 #status predicted <P15>

Alignment Scores:  
Pred. No.: 1.75e-105 Length: 500  
Score: 2233.00 Matches: 422  
Percent Similarity: 90.35% Conservative: 37  
Best Local Similarity: 83.07% Mismatches: 35  
Query Match: 78.63% Indels: 14  
DB: 1 Gaps: 5

US-09-475-704A-4 (1-1509) x FOVWH4 (1-500)  
QY 1 ATGGGCGCGCGCGCGCATCTCGCGCGGAGAGCTGGAACAAGTGGAGAAAGATCCGC 60  
Db 1 MetGIyAlaArgAlaSerValLeuSerGIyGIyGIuLeuAspArgTyrGIuLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGCAAGACACTACATGCTGAAGCACTGTGTGGCGCAGCGCGAG 120  
Db 21 LeuArgProGIyGIyLysLysGIInTyrArgLeuLysHisIleValTyrAlaSerArgLys 40  
QY 121 CTGAGGGCTTTCGCGCTGAACCCCGGCGCTGTGAGACCGCGCGAGGGCTGCAAGCAGATC 180  
Db 41 LeuGIuArgPheAlaValAsnProGIyLeuLeuGIuThrSerLysGIyCysArgGIInIle 60  
QY 181 ATGAAGCAGCTGACGCGCGCGCGCGCAAGCCGAGAGAGCTGCGAGCTGTATCAAC 240  
Db 61 LeuGIyGIuLeuGIInProSerLeuGIInThrGIySerGIuGIuLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCCTGTACTGCGTGCAAGCGCGCATCGAGGTCCGCAACCAAGAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGIuValArgAspThrLysGIuAla 100  
QY 301 CTGACCAAGATCGAGAGGAGCAGAAACAAGTCCAGCAGAAAGACCAGAGGCCAAGAG 360

```
Db 101 LeuAspIysIleGIuGIuGIuGlnAsnIysSerIySlySlySAlaGlnGlnAlaAla 120
QY 361 GCCGACGGC-----AAGGTGAGCCAGAACTACCCCATCGTGCAAGAACTGCAGGGC 411
Db 121 AspThrGIyAsnSerSerGIuValSerGIuAsnTyrProIleValGIuAsnLeuGIuGIy 140
QY 412 CAGATGGTGCAACGAGGCATCAGCCCCCGGCAACCTGAACGCTGGGTGAAGGTGATCGAG 471
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleValIySlyValIleGIu 160
QY 472 GAGAGGCTTCAGGCCCCGAGGTGATCCCATGTTACCGCCCTGAGGAGGGCGCCAC 531
Db 161 GIuIySAlaPheSerProGIuValIleProMetPheAlaAlaLeuSerGIuGIyAlaThr 180
QY 532 CCCCAGGACCTGAACACGATGTTGAACACCGTGGCGGCGCACGAGCCCATGACAGATG 591
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGIyGIyHisGlnAlaAlaMetGlnMet 200
QY 592 CTGAAGGACACCATCAACGAGGAGGCGCGAGTGGAGCCGCTGACCCCGTGCAGGCC 651
Db 201 LeuIySgluThrIleAsnGIuGIuAlaAlaGIuIleTyrAspArgLeuHisProValHisAla 220
QY 652 GGGCCCGTGGCCCCCGGACGAGATGCGCGACCCCGCGGACGACATCGCCGCGCCAC 711
Db 221 GlyProIleAlaProGIyGlnMetArgGIuProArgGIySerAspIleAlaGIyThrThr 240
QY 712 AGCACCTGACGAGACGATCGCTGATGACCAACCCCGCTGCGCGTGGCGGAC 771
Db 241 SerThrLeuGIuGIuGlnIleGIyTyrMetThrAsnAsnProProThrProValGIyGIu 260
QY 772 ATCTACAAGCGGTGATCATCTCGGGCTGAACAAGATCGTGGAGATGTACAGCCCGGTG 831
Db 261 IleTyrIySArgTyrIleIleLeuGIyLeuAsnIySleIleValArgMetTyrSerProIle 280
QY 832 AGCATCTGACATCCGCGACGAGGGCGCCCAAGAGGCCCTTCCGCACTACGTGACCGCTTC 891
Db 281 SerIleLeuAspIleArgGIuGIyProIySgluProPheArgAspTyrValAspArgPhe 300
QY 892 TTCAGAACCCTGCGCGCGGACGAGGCCACCCAGACGTCGAAGAACTGATGACCGAGACC 951
Db 301 TyrIySThrLeuArgAlaGIuGlnAlaSerGIuGIuValIySAsnTyrMetThrGIuThr 320
QY 952 CTGCTGCTGACGAAGCGCAACCCCGCATGCAAGACCATCTCGCGCTCTCGGCCCGGC 1011
Db 321 LeuLeuValGIuAsnAlaAsnProAspCysIySThrIleLeuIySAlaLeuGIyProAla 340
QY 1012 GCCACCTGAGAGAGATGATGACCGCTGCCAGGGCGTGGGGCGGCCGACCAAGGCC 1071
Db 341 AlaThrIleuGIuGIuMetMetThrAlaCysGIuGIyValGIyGIyProGIyHisIySAla 360
QY 1072 CGCGTGTGGCCGAGCGATGAGCCAG---GCCAACGCGTGAACATCATGATGCAGAA 1128
Db 361 ArgValIleuAlaGIuAlaMetSerGIuValThrAsnSerAlaThrIleMetMetGlnArg 380
QY 1129 AGCACTTCAAGGGCCCCCGCGCAAGTCAAGTCTTCACTGCGGCAAGAGGGCCAC 1188
Db 381 GIyAsnPheArgArgGIuGIyIySThrValIySlyCysPheAsnCysGIyIySgluGIyHis 400
QY 1189 ATCGCCAAGAACTGCGCGCGCCCCCGGCAAGAGGGCTGGAAGTCCGGCAAGAGGGCC 1248
Db 401 IleAlaArgAsnCysIySAlaProArgIySlySlyCysTyrIySlyCysGIyArgGIuGIy 420
QY 1249 CACCAAGTGAAGACTGCACCGAGCGGCAAGGCCAACTTCTGGGCAAGATCTGGCCCA 1308
Db 421 HisGlnMetIySAspCysThrGIuArgGIuAlaAsnPheLeuGIyIySleTyrProSer 440
QY 1309 CACAGGGCGCGCCCGCAACTTCTCTGACAGAACCGGACGAGCCCGCGCCCAACCGTG 1368
Db 441 HisIySgluIyArgProGIyAsnPheLeuGIuInsArgProGIu-----454
QY 1369 CCGACCGCGCGCGCGCGAGAGCTTCGCTTC-----GAGGAGACCAACCCCGCGCCCC 1422
Db 455 ProThrAlaProProGIuGIuSerPheArgPheGIyAspGIuThrThrThrProSerGIu 474
```

```
QY 1423 AAGCAGAGCCCAAGACCGGAGCCCTACCGGAGCCCTGACCGCCCTGCGACGCTG 1482
Db 475 IySgluGIuProArgAspIySgluLeuTyr-----ProLeuAlaSerLeuArgSerLeu 492
QY 1483 TTGGCAGCGCGCCCTGAGCCAG 1506
Db 493 PheGIyAsnAspProSerSerGIu 500

RESULT 5
S54377
gag polyprotein - human immunodeficiency virus type 1
C/Species: human immunodeficiency virus type 1, HIV-1
C/Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
R/Theodore, T.; Buckler-White, A.J.
submitted to the EMBL Data Library, July 1989
A/Reference number: S54377
A/Accession: S54377
A/Status: preliminary
A/Molecule type: genomic RNA
A/Residues: 1-501 <THE>
A/Cross-references: UNIPROT:P12495; EMBL:M22639; NID:g329377; PIDN:AAA45365.1; PID:g3293
C/Superfamily: AIDS-related virus gag polyprotein
C/Keywords: polyprotein

Alignment Scores:
Pred. No.: 1.85e-105 Length: 501
Score: 2232.50 Matches: 424
Percent Similarity: 90.37% Conservative: 36
Best Local Similarity: 83.30% Mismatches: 34
Query Match: 78.61% Indels: 15
DB: 2 Gaps: 5

US-09-475-704A-4 (1-1509) x S54377 (1-501)
QY 1 ATGGGCGCGCGCGCAGCATCTCGCGCGGCGAGAGCTGGAACAGTGGAGAAATCCGC 60
Db 1 MetGIyAlaArgAlaSerValLeuSerGIyGIyIySleuAspAlaIleProGIuIySleArg 20
QY 61 CTGGCGCCCGCGCGGAAGACATCATGCTGAAGCACCTGTGTGGCGGCGGAG 120
Db 21 LeuArgProGIyGIyIySlySlySlyTyrArgLeuIySlySlyValIleThrAlaSerArgGIu 40
QY 121 CTGAGGGCTTGGCCCTGAACCCCGGCTGTGAGACCGCGGAGGCTGCAAGCATC 180
Db 41 LeuGIuArgPheAlaLeuAsnProGIyLeuLeuGIuThrSerAspGIyCysIySgluIle 60
QY 181 ATGAAGCAGCTGCACCCCGGCTGCAGACCGGACCGAGGAGTCCGCGAGCTGTACAAC 240
Db 61 IleGIyGlnLeuGlnProAlaIleArgThrGIySerGIuIyLeuArgSerLeuPheAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGCGGCGCATCGAGGTCCGCGACACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGIuArgIleGIuValIySAspThrIySgluAla 100
QY 301 CTGCAACAAGATCCAGAGGAGGACCAACAAGTCCCAAGCAAGAACCAAGAGGAGAG 360
Db 101 LeuGIuIySmetGIuGIuGIuGlnAsnIySerySAsnIySlySAlaGlnGlnAlaAla 120
QY 361 GCCGAC-----GCCAAGTGAAGCCCAACTACCCCATCTGTGCAAGAACTGCGAG 408
Db 121 AlaAspAlaGIyAsnAsnSerGIuValSerGIuAsnTyrProIleValGIuAsnLeuGIu 140
QY 409 GGGCAGATGTGACCAAGGCCATCAGCCCCCGGCAACCTGAAACGCTGGGTGAAGGTGATC 468
Db 141 GIyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleProValIySValIle 160
QY 469 GAGGAGAAAGGCTTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCC 528
Db 161 GIuGIuIySAlaPheSerProGIuValIleProMetPheSerAlaLeuSerGIuGIyAla 180
QY 529 ACCCCCAAGACTGAACAAGATGTTGAACACCGTGGGCGGCCACAGGCCCATGCGAG 588
```

Db 181 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 200  
Qy 589 ATGCTGAAGACACCATCAACGAGAGGCGCCGAGTGAGACCGCTTGACCCCGTGAG 648  
Db 201 MetLeuysGlnThrIleAsnGlnGlnAlaAlaGlnTrpAspArgLeuHisProValHis 220  
Qy 649 GCCGGCCCGTGCGCCCGCCAGATGCGCGAGCCCGCGGAGCGACATCGCCGCGCC 708  
Db 221 AlaGlyProIleAlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGlyThr 240  
Qy 709 ACCAGCACCTTGAGAGAGATCGCTGATGACCAACCCCGCGTGCCCGTGAGC 768  
Db 241 ThrSerThrLeuGlnGlnGlnIleAlaTrpMetThrSerAsnProIleProValGly 260  
Qy 769 GACATCAACAAGCGGTGATCATCTGCGGCTGAACAAGATCGTGCGGATGTACAGCCC 828  
Db 261 GlnIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 280  
Qy 829 GTGAGCATCTGACATCCGCGAGGCGCCCAAGAGAGCCCTTCGCGACTACGTGAGCCG 888  
Db 281 ValSerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArg 300  
Qy 889 TTCTTCAAGACCTTGCGCGCCGAGAGGCGCACCCAGACGTGAAGAATGATGACCGAG 948  
Db 301 PheTyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysGlyTyrMetThrGln 320  
Qy 949 ACCCTGCTGTGACAGACGCGCAACCCGCACTGCAAGACCATCTGCGCGCTCTCGGCCC 1008  
Db 321 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 340  
Qy 1009 GGCGCCACCTTGAGAGATGATGACCGCTGCGAGGCGTGCGCGCCCGCGCAAG 1068  
Db 341 GlnAlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLys 360  
Qy 1069 GCCCGCGTGTGCGCGAGGCGGATGAGCCAGGCC-----AACAGCGTGAACATCATGATG 1122  
Db 361 AlaArgValLeuAlaGlnAlaMetSerGlnAlaThrAsnSerAlaAlaAlaValMetMet 380  
Qy 1123 CAGAAGAGCAACTTCAAGGGCGCCCGCGCAACGTCAAGTGTCTCAACTGCGCAAGAG 1182  
Db 381 GlnArgGlyAsnPheLysGlyProArgLysThrIleLysCysPheAsnCysGlyLysGln 400  
Qy 1183 GGCCACATCGCCCAAGAACTGCGCGCGCCCGCGCAAGAGGCTGTGGAAGTGCGGCAAG 1242  
Db 401 GlyHisIleAlaLysAsnCysArgAlaProArgArgLysGlyCysTrpLysCysGlyLys 420  
Qy 1243 GAGGCGCACCATGAAAGACTGACCGAGCGCGCAACTTCTTGCGCAAGATCTGG 1302  
Db 421 GlnGlyHisGlnLeuLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrp 440  
Qy 1303 CCCAGCCCAAGGGCGCGCCCGCGCAACTTCTTGCAAGACCGCGAGCGCGCGCGCCCC 1362  
Db 441 ProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 456  
Qy 1363 ACCGTGCCACCGCGCCCGCGCGAGAGCTTCCGCTTC---GAGAGAGACCAACCCCGCC 1419  
Db 457 -----ProThrAlaProProAlaGlnSerPheGlyPheGlyGlnGlnIleThrProSer 474  
Qy 1420 CCCAAGCAGAGCCCAAGAGACCGCGAGCGCCCTTACCGCGAGCCCTGACCGCGCGAGC 1479  
Db 475 GlnLysGlnGlnGlnLysAspLysGlnLeuTyr-----ProSerThrAlaLeuLysSer 492  
Qy 1480 CTGTTGCGGAGCGCGCCCGCTGAGCCAG 1506  
Db 493 LeuPheGlyAsnAspProLeuLeuGln 501

RESULT 6  
FOVWLV

gag polypeptide - human immunodeficiency virus type 1 (isolate LAV-1a)  
N:Alternate names: assemblin; core polypeptide; gag precursor  
N:Contains: capsid antigen core protein p24CA; core protein p1; core protein p6; matrix  
C:Species: human immunodeficiency virus type 1, HIV-1

A/Note: host Homo sapiens (man)  
C/Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 05-Dec-1998  
C/Accession: A03946  
R/Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.  
Cell 40, 9-17, 1985  
A/Title: Nucleotide sequence of the AIDS virus, LAV.  
A/Reference number: A90866; MUID:8509333; PMID:2981635  
A/Accession: A03946  
A/Molecule type: DNA  
A/Residues: 1-500 <WAI>  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polypeptide  
C/Keywords: AIDS; blocked amino end; core protein; immunodeficiency; lipoprotein; myrist  
F/2-500/Product: gag precursor (assemblin) #status predicted <GAG>  
F/2-132/Product: matrix antigen core protein p17MA #status predicted <P17>  
F/20-32/Region: nuclear location signal  
F/110-114/Region: nuclear location signal  
F/133-363/Product: capsid antigen core protein p24CA #status predicted <P24>  
F/364-377/Product: core protein p2 #status predicted <CP2>  
F/378-432/Product: nucleocapsid core protein p7NC #status predicted <CP7>  
F/392-405/Region: zinc finger CCHC motif  
F/413-426/Region: zinc finger CCHC motif  
F/433-448/Product: core protein p1 #status predicted <CP1>  
F/449-500/Product: core protein p6 #status predicted <CP6>  
F/2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted  
F/392,395,400,405/Binding site: zinc (Cys, Cys, His, Cys) #status experimental  
F/413,416,421,426/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

Alignment Scores:

Pred. No.: 2.47e-105 Length: 500  
Score: 2230.00 Matches: 421  
Percent Similarity: 90.55% Conservative: 39  
Best Local Similarity: 82.87% Mismatches: 34  
Query Match: 78.52% Indels: 14  
DB: 1 Gaps: 5

US-09-475-704A-4 (1-1509) x FOVWLV (1-500)

Qy 1 ATGGGCGCGCGCGCGAGCATCTGCGCGCGAGAGCTGGACAACTGGGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGlnLeuAspArgTrpGlnLysIleArg 20  
Qy 61 CTGCGCGCGCGCGCGCAAGAACTACTGCTGAAGACCTGTGTGGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
Qy 121 CTGAGGCGCTTCCGCTGAACCCCGCGCTGTGAGACCGCGAGGCGCTGCAAGACATC 180  
Db 41 LeuGlnArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60  
Qy 181 ATGAAGCAGCTGACGCGCGCGCGCTGCAACCGCGAGAGAGAGAGCTGCGGAGCTGTACAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnLeuArgSerLeuTyrAsn 80  
Qy 241 ACCGTGCCACCTGTGACTGCTGCGTGCACGCGCGCATGAGAGTCCGCGACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGlnIleLysAspThrLysGlnAla 100  
Qy 301 CTGAGCAGATGAGAGAGAGAGAGACAACTGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120  
Qy 361 GCCGAGCGC-----AAGTGAGCGAGACTACCCCATCTGTGAGAGAGAGAGAGAGAG 411  
Db 121 AspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly 140  
Qy 412 CAGATGTCACACGAGGACATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValGln 160  
Qy 472 GAGAAGCCTTCAAGCGCGCGAGGTGATCCCATGTTCACCGCGCTGAGAGAGAGAGAGAG 531

```
Db      161 GlulysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180
QY      532 CCCGAGGACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCGCCATGCAATG 591
Db      181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMet 200
QY      592 CTGAAGACACCATCAACGAGGAGCGCCGAGTGGACCGCTGCACCCCGTGCAGGCC 651
Db      201 LeuLysGluThrIleAsnGluGluAlaAlaGluThrAspArgValHisProValHisAla 220
QY      652 GGGCCCGTGGCCCCCGGCGCAGATGCGGCACCCCGCGCGCAGCAGCATGCGCGCGCAC 711
Db      221 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 240
QY      712 AGCACCTTGAGAGAGAGATCGCTGGATGACCAACCCCGCTGCGCGCTGGCGGAC 771
Db      241 SerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsnProProIleProValGlyGlu 260
QY      772 ATCTACAAAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTG 831
Db      261 IleTyLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTySerProThr 280
QY      832 AGCATCCTGAGACATCCGCCAGGCGGCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTT 891
Db      281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300
QY      892 TTCAGAAGCCCTGCGCGCGCAGAGGCCACCCAGACGCTGAAGAACTGGATGACCGAGACC 951
Db      301 TyLysThrLeuArgAlaGluGlnAlaSerGlnValLysAsnTrpMetThrGluThr 320
QY      952 CTGCTGTGTGAGAAACGCCAACCCCGCATGTCAGAACCATCTCGCGCTTCCGCGCGGC 1011
Db      321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340
QY      1012 GCCACCTTGAGAGAGATGATGACCGCTGCCAGGCGGTGGCGCGCGCCGACCAAGGCC 1071
Db      341 AlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla 360
QY      1072 CGCGTGTGGCCGAGCGATGAGCCAG--GCCAACGCTGAACATCATGATGACGAAG 1128
Db      361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArg 380
QY      1129 AGCAACTTCAAGGGCCCCGCGCGCAACGTCAAGTCTTCAACTGCGCGCAAGAGGGCCAC 1188
Db      381 GlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCysGlyLysGluGlyHis 400
QY      1189 ATGCCCAAGAACTGCGCGCGCGCGCGCAAGAGGGCTGCTGAAGTGGCGCAAGAGGGCC 1248
Db      401 IleAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGly 420
QY      1249 CACCAGATGAAGGACTGCACCGAGCGCAGGCCAATCTCTGGCAAGATCTGGCCAGC 1308
Db      421 HisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440
QY      1309 CACAAGGGCGCGCGCGCAACTTCTGCAGAGAACCGCAGCGAGCCCGCGCCCAACCGTG 1368
Db      441 TyLysGlyArgProGlyLysAsnPheLeuGlnSerArgProGlu----- 454
QY      1369 CCCACCGCGCGCGCGCGAGAGCTTCCG-----TTCGAGGAGACCAACCCCGCGCCCC 1422
Db      455 ProThrAlaProProGluGluSerPheArgSerGlyValGluThrThrThrProSerGln 474
QY      1423 AAGCAGAGAGCCCAAGAGACCGCGAGCCCTACCGCGCCCTGACCGCCCTGCGCAGCCTG 1482
Db      475 LysGlnGluProIleAspLysGluLeuTy-----ProLeuThrSerLeuArgSerLeu 492
QY      1483 TTCGGCAGCGCGCGCGCTGAGCCAG 1506
Db      493 PheGlyAsnAspProSerSerGln 500

RESULT 7
A44001
gag polyprotein - human immunodeficiency virus type 1 (strain YU-2)
```

```
N/Alternate names: core polyprotein
N/Contains: core protein p15; core protein p17; core protein p24
C/Species: human immunodeficiency virus type 1, HIV-1
A/Note: host Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C/Accession: A44001
R/Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.
J. Virol. 66, 6587-6600, 1992
A/Title: Complete nucleotide sequence, genome organization, and biological properties of
A/Reference number: A44001; MUID:93021387; PMID:1404605
A/Accession: A44001
A/Molecule type: DNA
A/Residues: 1-500 <LiY>
A/Cross-references: UNIPROT:P35962; GB:M93258
C/Genetics:
A/Gene: gag
C/Superfamily: AIDS-related virus gag polyprotein
C/Keywords: AIDS; core protein; immunodeficiency; polyprotein
F/1-132/Product: core protein p17 #status predicted <P17>
F/133-391/Product: core protein p24 #status predicted <P24>
F/392-500/Product: core protein p15 #status predicted <P15>

Alignment Scores:
Pred. No.: 3,51e-105 Length: 500
Score: 2227.00 Matches: 423
Percent Similarity: 89.96% Conservative: 34
Best Local Similarity: 83.27% Mismatches: 37
Query Match: 78.42% Indels: 14
DB: 1 Gaps: 5

US-09-475-704A-4 (1-1509) x A44001 (1-500)
QY      1 ATGGGCGCGCGCGCGCAGCATCTCGCGCGGAGAGAGCTGGAACAAGTGGAGAAATCCGC 60
Db      1 MercGlyAlaArgAlaSerValLeuSerAlaGlyGluLeuAspLysTrpGluLysIleArg 20
QY      61 CTGCGCGCGCGCGCGCAAGACATACATGCTGGAAGCACCTGTGTGGCCAGCGCGAG 120
Db      21 LeuArgProGlyGlyLysLysGlnTyArgLeuLysHisIleValThrAlaSerArgGlu 40
QY      121 CTGAGGCGCTTGCCTGAACCCCGCGCTGTGAGACCGCGCGAGCGCTGCAAGCAGATC 180
Db      41 LeuGluArgPheAlaValAlaAspProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60
QY      181 ATGAAGCAGCTGAGCGCGCGCTGCAGACCGGCAACGAGAGCTGCGCAGCTGTACAAC 240
Db      61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80
QY      241 ACCGTGGCCACCTGTACTGCGTGCACGCGCGCATCGAGTCCGCGACACCAAGAGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHisGlnLysIleGluValLysAspThrLysGluAla 100
QY      301 CTGGAACAAGATGAGAGAGAGAGCAACAAGTCCACGAGAACCCAGACGCGCAAGAG 360
Db      101 LeuGluLysIleGluGluGluGlnAsnLysSerLysLysValGlnGlnAlaAlaAla 120
QY      361 GCCGACGCGC-----AAGGTGAGCCAGAACTAACCCTATCGTGCAAGACCTGAGGGC 411
Db      121 AspThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140
QY      412 CAGATGTGTACACGAGCCATCAACCCCGCACCTTGAACGCGCTGGTGAAGTATCGAG 471
Db      141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleTrpValLysValGlu 160
QY      472 GAGAAGGCGCTTACGCGCGGAGGTGATCCCAATTTCAACCGCGCTGAGCGAGGGGCCACC 531
Db      161 GlulysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180
QY      532 CCCGAGGACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCGCCATGCAATG 591
Db      181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMet 200
QY      592 CTGAAGGACACCATCAACGAGGAGCGCGCGAGTGGAGCCGCGCTGCACCCCGTGCAGGCC 651
```

```
Db      201 LeuylsGluThrIleAsnGluGluAlaIleGluThrAspArgLeuHisProValHisAla 220
Qy      652 GGGCCCGTGGCCCCCGGCGAGATGCGCGACCCCGCGGCGAGCGACATCGCCGCGCCACC 711
Db      221 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleIleGlyThrThr 240
Qy      712 AGCACCTGTGAGAGAGACATCGCTGTGATGACCAACCCCCCGTGCCTGGGCGAC 771
Db      241 SerThrLeuGlnGlnGlnIleGlyTyrMetThrAsnAsnProIleProValGlyGlu 260
Qy      772 ATCTACAAGCGGTGATCATCTGGGCTGAACAAGATGTCGATGATACAGCCCGTG 831
Db      261 IleTyrLysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280
Qy      832 AGCATCTGTGACATCCGCGGAGGCCCCCAAGAGAGCCCTTCGCGACTAGTGGACCGCTTC 891
Db      281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300
Qy      892 TTCAGACCTGTGCGCGCGGAGAGCGCAACCCAGACGTAAGAAGACTGATGACCGAGACC 951
Db      301 TyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTyrMetThrGluThr 320
Qy      952 CTGCTGTGTGAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGCGCCGCC 1011
Db      321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340
Qy      1012 GCCACCTGTGAGAGATGATGACCGCTGCCAGGGCGTGGGCGCGCCCGGCGCAAGGCC 1071
Db      341 AlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisValAla 360
Qy      1072 CGCGTGTGCGCGGAGCGGATGAGCCAG--GCCAACAGCGTGAAACATGATGAGAGAG 1128
Db      361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArg 380
Qy      1129 AGCAACTTCAAGGGCCCCCGGCGCAAGCTCAAGTCTTCAACTGCGGCAAGAGGGCCAC 1188
Db      381 GlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluLysHis 400
Qy      1189 ATGCGCAAGACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGAGGC 1248
Db      401 IleAlaLysAsnCysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGluGly 420
Qy      1249 CACCAAGATGAAGACTGCACCGGAGCGCGCAAGGCTGTGAAAGTGGCGCAAGTGGCCAGC 1308
Db      421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTyrProSer 440
Qy      1309 CACAAGGGCGCGCGCGCAACTTCTGTGCAGAACCGGACGAGCGCGCGCCCGCACCGTG 1368
Db      441 HisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 454
Qy      1369 CCCACCGCGCGCGCGCGAGAGCTTCGCTTC-----GAGAGAGACCAACCCCGCGCCC 1422
Db      455 ProThrAlaProSerGlnGlnSerValArgPheGlyGlnGluThrThrProSerGln 474
Qy      1423 AAGCAGAGAGCCCAAGAGACCGGAGCCCTTACCGGAGCCCTGACCGCGCTGGCGAGCCTG 1482
Db      475 LysGlnGlnProIleAspLysGlnLeuTy-----ProLeuAlaSerLeuArgSerLeu 492
Qy      1483 TTCGGCAGCGCGCGCGCTGAGCCAG 1506
Db      493 PheGlySerAspProSerSerGln 500

RESULT 8
S33979
gag polyprotein - human immunodeficiency virus type 1
C/Species: human immunodeficiency virus type 1, HIV-1
C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C/Accession: S33979
R/Carlini, F.
submitted to the EMBL Data Library, November 1991
A/Reference number: S33979
A/Accession: S33979
```

```
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-500 <CAR>
A/Cross-references: UNIPROT:Q78240, EMBL:Z11530, NID:g60192, PIDN:CAA77621.1, PID:g60193
C/Superfamily: AIDS-related virus gag polyprotein

Alignment Scores:
Pred. No.: 3,51e-105 Length: 500
Score: 2227.00 Matches: 420
Percent Similarity: 90.16% Conservative: 38
Best Local Similarity: 82.68% Mismatches: 36
Query Match: 78.42% Indels: 14
DB: 2 Gaps: 5

US-09-475-704A-4 (1-1509) x S33979 (1-500)
Qy      1 ATGGGCGCGCGCGCCAGCATCTGCGCGGCGAGAGCTGGAACAAGTGGAGAGATCCGC 60
Db      1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLysLeuAspArgTyrGluLysIleArg 20
Qy      61 CTGCGCGCGCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCCAGCGCGAG 120
Db      21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTyrAlaSerArgGlu 40
Qy      121 CTGAGAGGCTTGCCTGAACCCCGCGCTGTGAGAGACCGCGGAGGCTGCAAGCAGATC 180
Db      41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluArgArgSerLeuTyrAsn 60
Qy      181 ATGAAGCAGCTGACGCGCGCTGACAGCGGCGACCGAGAGAGTGGCGAGCTGTACAC 240
Db      61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluArgArgSerLeuTyrAsn 80
Qy      241 ACCGTGGCCACCTGTACTGCTGACAGCGCGGATCGAGGTCCGCGAGCAACCAAGAGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100
Qy      301 CTGACAAAGATCGAGAGAGAGAGCAAGTCCCGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db      101 LeuAspLysIleGluGluGlnGlnAsnLysSerLysLysAlaGlnGlnAlaAla 120
Qy      361 GCCGAGCGC-----AAGTGAGCGAGAACTACCCCATGCTGTGAGAACTGAGAGGC 411
Db      121 AspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly 140
Qy      412 CAGATGTGCACAGAGGCATCAGCCCGCGACCCCTGAACGCTGGGTGAAGTGAAGAG 471
Db      141 GlnMetValHisGlnAlaLeuSerProArgThrLeuAsnAlaTyrValLysValGlu 160
Qy      472 GAGAAGGCTTCAAGCGCGAGGTGATCCCATGTTCAACCGCGCTGAGAGAGCGCGCAC 531
Db      161 GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180
Qy      532 CCCCAGACCTGAACAGATGTTGAACACCGTGGCGCGCACAGCGCGCATGAGATG 591
Db      181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMet 200
Qy      592 CTGAAGACACCATCAACGAGAGCGCGGAGTGGAGACCGCTGCAACCCGTCGAGAGCC 651
Db      201 LeuLysGluThrIleAsnGlnGluAlaIleGluTyrAspArgLeuHisProValGlnAla 220
Qy      652 GGGCCCGTGGCCCCCGGCGAGATGCGCGACCCCGCGGAGAGATGCGCGCGCAC 711
Db      221 GlyProIleAlaProGlyGlnIleArgGluProArgGlySerAspIleAlaGlyThrThr 240
Qy      712 AGCACCTGTGAGAGACATCGCTGTGATGACCAACCCCGTGCCTGGGCGAG 771
Db      241 SerThrLeuGlnGlnGlnIleArgTyrMetThrAsnAsnProIleProValGlyGlu 260
Qy      772 ATCTACAAGCGGTGATCATCTGGGCTGAACAAGATGTCGATGATGATACAGCCCGTG 831
Db      261 IleTyrLysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280
Qy      832 AGCATCTGTGACATCCGCGGAGGCCCCCAAGAGAGCCCTTCGCGACTAGTGGACCGCTTC 891
```

Db 281 SerIleuAspIleArgGlnGlyProIySgluProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAAGACCCCTGGCGCGGAGCAGGCCACCCAGACGCTGAAGAAGACTGATGACCGAGACC 951  
Db 301 TyrIleThrIleuArgIaGlnIaMetSerGlnIuValIySAsnTrpMetThrGluThr 320  
QY 952 CTGCTGTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCGCC 1011  
Db 321 LeuIeuValGlnAsnIaAsnProAspCysIySThrIleuIySAlaIeuGlyProGlu 340  
QY 1012 GCCACCTTGAGAGAGATGATGACCGCTGCGCAGGGCGTGGCGCGCCCGCCACAGGCC 1071  
Db 341 AlaThrIeuGlnIuIuMetMetThrAlaCysGlnIyValGlyIyProGlyHisIySAla 360  
QY 1072 CGCGTGTGGCGGAGGCGCATGAGCCAG---GCCAACAGCGTGAACATCATGATGACAGAG 1128  
Db 361 ArgValIleuAlaGlnIaIaMetSerGlnIValThrAsnProAlaAsnIleMetMetGlnArg 380  
QY 1129 AGCAACTTCAAGGGCGCGCGCGCAACGTCAGTCTTCAACTGCGGCAAGAGAGGCCAC 1188  
Db 381 GlyIySPhaArgAsnGlnIyIySThrValIySPhaAsnCysGlyIySgluGlyHis 400  
QY 1189 ATCGCCAAAGACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGTGCGCAAGAGGCC 1248  
Db 401 IleAlaArgAsnCysArgAlaProArgIySgluIySlyCysTrpIySlySgluGly 420  
QY 1249 CACCAGATGAAGACTGCAACCGAGCGCCACCACTTCTGGGCAAGATCTGCGCCAGC 1308  
Db 421 HisGlnMetIySAspCysThrGlnArgGlnAlaAsnPhaIeuGlyIySleTrpProSer 440  
QY 1309 CACAAGGGCGCGCGCGCAACTTCTTGAGAACCGCAGCGAGCGCGCGCGCCACCGCTG 1368  
Db 441 HisIySgluArgProGlyAsnPhaIeuGlnSerArgProGlu----- 454  
QY 1369 CCCACCGCGCGCGCGCGAGAGCTTCCGCTTC-----GAGGAGACCAACCCCGCGCCC 1422  
Db 455 ProThrIaProProGluGluIuSerPheArgPheGlyGluIuThrThrProSerGln 474  
QY 1423 AAGCAGAGACCCCAAGAGCCGAGCCCTTAACCGCGAGCCCTGACCGCGCTGCGAGCGCTG 1482  
Db 475 LysGlnIuProIleAspIySgluMetTyr-----ProIeuAlaSerIeuArgSerIeu 492  
QY 1483 TTCGGCAGCGCGCGCGCTGAGCCAG 1506  
Db 493 PheGlyAsnAspProSerSerGln 500

RESULT 9

gag polypeptide - human immunodeficiency virus type 1  
C;Species: human immunodeficiency virus type 1, HIV-1  
C;Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 09-Jul-2004  
C;Accession: T01667  
R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.  
Cell 46, 63-74, 1986  
A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol  
A;Reference number: Z14389; MUID:86245056; PMID:2424612  
A;Accession: T01667  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-505 <All>  
A;Cross-references: UNIPROT:P04594; EMBL:K03456; MID:g60228; PIDN:CAA28011.1; PID:g60229  
C;Superfamily: AIDS-related virus gag polypeptide

Alignment Scores:

Pred. No.: 4.78e-104 Length: 505  
Score: 2204.50 Matches: 422  
Percent Similarity: 88.33% Conservative: 32  
Best Local Similarity: 82.10% Mismatches: 39  
Query Match: 77.62% Indels: 21  
DB: 2 Gaps: 6

US-09-475-704A-4 (1-1509) x T01667 (1-505)

QY 1 ATGGGCGCGCGCGGACATCTCTGCGCGGAGAGCTGACAAGTGGGAGAAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValIeuSerGlyGlyIySleuAspAlaTrpGluIySleArg 20  
QY 61 CTGCGCGCGCGCGGCAAGACACTATGCTGAAGACCTGTGTGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyIySlySlySlySlySlySlySlySlySlySlySlySlySlySly 40  
QY 121 CTGAGAGGCTTCCGCTGAACCCCGCGCTGAGAGACCGCGAGGCTGCAAGACATC 180  
Db 41 LeuGluArgPheAlaIeuAsnProGlyIleuIeuGluThrGlyGluIyCysGlnIle 60  
QY 181 ATGAGCAGCTGCAGCGCGCGCGCGCGACCGGACGAGAGCTGCCAGCTGTACAC 240  
Db 61 MetGluIuIeuGlnSerThrIeuIySThrGlySerGluIuIleIySleuIySAsn 80  
QY 241 ACCGTGGCCACCCCTGTACTGCGTGAACCGCGGCAATGAGTCCGCGCACCAAGAGGCC 300  
Db 81 ThrValAlaThrIeuIySValHisGlnArgIleAspValIySAspThrIySgluAla 100  
QY 301 CTGCAACAAGTGCAGAGAGAGACAAAGTCCCAACAGAGACCAAGAGAGAG 360  
Db 101 LeuAspIySleGluIuIleGlnAsnIySleArgGlnIySThrGlnIuAlaAla 120  
QY 361 GCC-----GACGCGAAGTGAAGCCAGAACTACCCCATC 393  
Db 121 AlaGlnIuAlaAlaAlaAlaThrIySAsnSerSerValSerGlnAsnTrpIle 140  
QY 394 GTGCAGAACCTGCAGAGGCGCGAGATGTCACAGGCGCATGACCCCGCGCACCTGAAGCC 453  
Db 141 ValGlnAsnAlaGlnIyGlnMetIleHisGlnAlaIleSerProArgThrIeuAsnAla 160  
QY 454 TGGGTGAAGTGAATGAGAGAGAGGCTTACGCGCGAGGTGATCCCATGTTACCGCC 513  
Db 161 TrpValIySValIleGluIuIySAlaPheSerProGluValIleProMetPheSerAla 180  
QY 514 CTGAGCGAGGCGCGCACCCCGCGAGACCTGAACAGATGTTGAACACCGTGGCGGCGAC 573  
Db 181 LeuSerGluIyAlaThrProGlnAspIeuAsnMetIeuAsnIleValGlyGlyHis 200  
QY 574 CAGCGCGCGCATGAGATGCTGAAGACACCATCAAGAGAGGCGCGAGTGGGAGCCGC 633  
Db 201 GlnAlaAlaMetGlnMetIeuIySAspThrIleAsnGluIuAlaAlaAspTrpAspArg 220  
QY 634 CTGCACCCCGTGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 693  
Db 221 ValHisProValHisAlaGlyProIleProProGlyGlnMetArgIuProArgIySer 240  
QY 694 GACATGCGCGCGCGCGCACCGACCGCTGAGAGAGATGCTGTGATGACCAAGACCCC 753  
Db 241 AspIleAlaGlyThrThrSerThrIeuGlnGlnIleGlyTrpMetThrSerAsnPro 260  
QY 754 CCCGTGCGCGTGGCGGACATCTTCAAGCGGTGATCATCTGGCGCTGAACAAGATGCTG 813  
Db 261 ProIleProValGlyAspIleTyrIySArgTrpIleIleIeuGlyIeuAsnIySleVal 280  
QY 814 CGGATGTACAGCGCGCTGAGCATCTGACATCGCGCAGGCGCCCAAGAGCGCTTCCGC 873  
Db 281 ArgMetTyrSerProValSerIleIeuAspIleArgGlnIyProIySgluProPheArg 300  
QY 874 GACTAGTGAACCGCTTCTTCAAGACCTGCGCGCGAGAGGCGCACCAAGACGTGAAG 933  
Db 301 AspTyrValAspArgPhePheIySThrIeuArgAlaGluGlnAlaThrGlnIuValIyS 320  
QY 934 AACTGATGACCGAGACCTGTGCTGTGAGAAAGCCCAACCCGCACTGCAAGACCATCTG 993  
Db 321 AsnTrpMetThrGluThrIeuIeuValGlnAsnAlaAsnProAspCysIySThrIleIeu 340  
QY 994 CGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1053  
Db 341 LysAlaIeuGlyProGlyAlaThrIeuGluIuIuMetMetThrAlaCysGlnIyValGly 360

DQ  
1054 GGGCCCCGCGCACAAGGCCCGCTGTGCCTCCGAGGCCATGAGCCAGCC-----AACAGC 1107

DB  
361 GlyProSerHisIlysaAlaArgValLeuAlaGluAlaMetSerGlnAlaThrAsnSerThr 380

DY  
1108 GTGAACATCATGATGTCAGAAGCAACTTCAAGGGCCCCGCGSCCAAGCTAAGTGCTTC 1167

DB  
381 AlaAlaIleMetMetGlnArgGlyAsnPhelysGlyGlnLysArg--IleIysCysPhe 399

DY  
1168 AACTGCGGCAAGAGGGGCCACATCGCCCAAGAACTGCGCGCGCCCCGCAAGAAGGCTGC 1227

DB  
400 AsnCysGlyLysGlyGlnIylHisLeuAlaArgAsnCysAlaGalaProAlaGlyIysGlyCys 419

DY  
1228 TGGAAGTGCGGCAAGAGAGGCCACCAGATGAAGGACTGCACCCAGCGCCAAGCCACTTC 1287

DB  
420 TrpLysCysGlyLysGlnIylHisGlnMetLysAspCysThrGluArgGlnAlaAsnPhe 439

DY  
1288 CTGGCAAGATCTGGCCCCAGCCACAAGGGCGCCCCCGCACTTCTCTGCAGAACCGCAGC 1347

DB  
440 LeuGlyLysIleTrpProSerHisIylsGlyArgProGlyAsnPhel euGlnInserArgPro 459

DY  
1348 GAGCCCGCGCGCCCCCACCGTGCACCGCCCCCGCGGAGAGCTTCCGCTTC--GAG 1404

DB  
460 Glu-----ProThrAlaProProAlaGluSerPheGlyPheGlyGln 473

DY  
1405 GAGACCAACCCCGCCCCCAAGCAGAGAGCCCAAGAGACCGCGAGCCCTACCGGAGCCCCTG 1464

DB  
474 GluIleLysPProSerGlnLysGlnGlnIylsAspLysGlnLeuTyr-----ProLeu 491

DY  
1465 ACCGCGCTGCGCAGCGCTGTTCGCGAGCGGCGCCCCCTGAGCCAG 1506

DB  
492 AlaSerLeuLysSerLeuPheGlyAsnAspGlnLeuSerGln 505

```

RESULT 10
T09436
gag polyprotein - human immunodeficiency virus type 1 (strain JRFL)
C/Species: human immunodeficiency virus type 1, HIV-1
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C/Accession: T09436
R/Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A/Reference number: Z16673
A/Accession: T09436
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-498 <PAN>
A/Cross-references: UNIPROT:Q75754; EMBL:U63632; NID:g1465777; PID:g1465778
C/Genetics:
A/Gene: gag
C/Superfamily: AIDS-related virus gag polyprotein

Alignment Scores:
Pred. No.: 1.02e-103 Length: 498
Score: 2198.00 Matches: 418
Percent Similarity: 90.32% Conservative: 39
Best Local Similarity: 82.61% Mismatches: 37
Query Match: 77.39% Indels: 12
DB: 2 Gaps: 6

```

US-09-475-704A-4 (1-1509) x T09436 (1-498)

[illegible]

181 ATGAAGCAGCTGCAGCCCGCCCTGCAGACCGGCACCGAGGAGCTGGCGAGCCTGTACAAC 240

Db	61	LeuGIgInLeuLeuProAlaLeuLysThrGIySerGIuLeuLysSerLeuTYrAsn	80
QY	241	ACCGTGGCCACCCCTGTACTGCGTGCAAGCCGGCATGCAAGTCCGCCACAACCAAGAGGCC	300
Db	81	ThrValAlaThrLeuTYrCYsValHisGlnArgTIGluValLysAspThrLysGluAla	100
QY	301	CTGCACAAGATCGAGAGGAGCAGACAAGTCCCGACAGAAAGACCAAGCCAGGCCAAGAG	360
Db	101	LeuGIuLysSILeGIuGIuGIuGlnAlaLysSerLYsLYsGluAlaAlaAlaAspThr	120
QY	361	GCCGAC---GGCAAGGTGAGCCAGAACTAACCCCATCGTCAGAACCTGCAGGGCCAGATG	417
Db	121	GIyAsnSerSerGIuValSerGIuAsnTYrProILeValGIuAsnMetGInGIyGlnMet	140
QY	418	GTGCACGAGGCCATCAGCCCCCGCACCCCTGAACGCCCTGGGTGAAGGTGATCGAGAGAA	477
Db	141	ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLYsValValGIuGluLys	160
QY	478	GCCTTCAGCCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGGGCCACCCCCCAG	537
Db	161	AlaPheSerProGIuValIleProMetPheSerAlaLeuSerGIuAlaIleThrProGln	180
QY	538	GACCTGAACACGATGTTGAACAACCGTGGCGGCGCCACAGGCCCGCATGCAATGCTGAAG	597
Db	181	AspLeuAsnThrMetLeuAsnThrValGIyGlnHisGlnAlaIleAlaMetGInMetLeuLys	200
QY	598	GACACCATCAACGAGGAGGCCCGCGAGTGGAGCGGCTGCACCCCGTGACAGCGCGCCCC	657
Db	201	GIuThrTrIeAsnGIuGluAlaIleGluTrpAspArgLeuHisProValHisAlaGIyPro	220
QY	658	GTGGCCCCCGCGCAGATGCGCGCAACCCCGCGCGAGCATCGCCGCGGCCACAGCAC	717
Db	221	ILeAlaProGIyGlnMetArgGIuProArgGIySerAspTrIleAlaGIyThrTrIleSerThr	240
QY	718	CTGCAGGACAGATCGCTGATGACCAAGCAACCCCCCGTGGCCCTGGCGCGCATCTAC	777
Db	241	LeuGIuGIuGIuGlnIleGIyTrpMetThrAsnAsnProProIleProValGIyGluIleTYr	260
QY	778	AAGCGGTGATCATCTTGCGCTGAACAAGATCGTGCGGATGTAACAGCCCCGTGAGCATC	837
Db	261	LysArgTrpTrIleIleLeuGIyLeuAsnLysIleValArgMetTYrSerProThrSerIle	280
QY	838	CTGCACATCCGCCAGGAGCCCAAGAGAGCCCTTCCGCGACTAGCTGAGCCGCTTCTCAAG	897
Db	281	LeuAspTrIleArgGIuGIyProLYsGIuProPheArgAspTYrValAspArgPheTYrLys	300
QY	898	ACCCTGCGCGCGAGCAGAGCCCAAGAGCGTGAAGAACTGGATGACCGGACCCCTGCTG	957
Db	301	ThrLeuArgIaGIuGlnAlaSerGIuValLYsAsnTrpMetThrGIuThrLeuLeu	320
QY	958	GTGCAGAAGCCCAACCCCGACTGCAAGACCACTCTGCGCGCTCTGGCCCGCGGCCAAC	1017
Db	321	ValGIuAsnAlaAsnProAspCYsLYsThrIleLeuLYsAlaLeuGIyProAlaAlaThr	340
QY	1018	CTGCAGGAGATGAGCCGCTGCGCAGAGCGGTGGCGCGCGCGGCCACAAGGCCCGCGTG	1077
Db	341	LeuGIuGIuMetMetThrAlaCYsGIuGlnValGIyGIyProGIyHisLYsValaArgVal	360
QY	1078	CTGGCCGAGCGGATGAGCCAG--GCCAACAGCGTGAACATCATGATGACAGAAAGCAAC	1134
Db	361	LeuAlaGIuAlaMetSerGIuValThrAsnProAlaThrIleMetMetGIuArgGIyAsn	380
QY	1135	TTCAGAGGCCCCCGCGCAAGCGTCAAGTGTCTCAACTGCGGCAAGAGGCGCACATGCGC	1194
Db	381	PheArgAsnGIuArgLYsIleValLYsCYsPheAsnCYsGIyLYsGIuGlnHisIleAla	400
QY	1195	AAGAACTGCGCGCGCCCCCGCAAGAGGCGTGTGAAGTGCAGAGGAGGCCACAG	1254
Db	401	ArgAsnCYsArgAlaProArgLYsLYsGIyCYsTYrLYsCYsGIyLYsGIuGlnHisGln	420
QY	1255	ATGAAGACTGCACCGAGCGCCAGGCAACTTCTCTGGCAAGATCTGGCCAGGCCACAAG	1314

100

DB 421 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerTyrLys 440  
QY 1315 GGGCGCCCCGGCACTTCTGAGACCGCAGCGCGCCGCCCAACCGTCCCAACC 1374  
DB 441 GlyArgProGlyAsnPheLeuGlnSerArgProGln-----ProThr 454  
QY 1375 GCGCGCCCCCGGAGAGCTTCCGCTTC--GAGGAGACC--ACCGCGCCCAAGCAG 1428  
DB 455 AlaProProGluLysSerPheArgPheGlyGluGlnThrAlaThrProSerGlnLysGln 474  
QY 1429 GAGCCCAAGACCGCGAGCCCTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTCGGC 1488  
DB 475 GluProIleAspLysGluMetTyr-----ProLeuThrSerLeuArgSerLeuPheGly 492  
QY 1489 AGCGGCCCCCTGAGCCAG 1506  
DB 493 AsnAspProSerSerGln 498

RESULT 11

A38068  
gag polyprotein - human immunodeficiency virus type 1 (strain MN)  
N/Alternate names: core polyprotein  
N/Contains: core protein p1; core protein p17; core protein p2; core protein p24; core p  
C/Species: human immunodeficiency virus type 1, HIV-1  
A/Note: host Homo sapiens (man)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C/Accession: A38068  
R/Henderson, L.B.; Bowers, M.A.; Sowder II, R.C.; Serabyn, S.A.; Johnson, D.G.; Bess Jr.  
J. Virol. 66, 1856-1865, 1992  
A/Title: Gag proteins of the highly replicative MN strain of human immunodeficiency viru  
A/Reference number: A38068; MUID:92194415; PMID:1548743  
A/Accession: A38068  
A/Molecule type: protein  
A/Residues: 1-506 <HEN>  
A/Cross-references: UNIPROT:P05888  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polyprotein  
C/Keywords: AIDS; core protein; immunodeficiency; polyprotein  
F/1-134/Product: core protein p17 #status experimental <P17>  
F/135-365/Product: core protein p24 #status experimental <P24>  
F/366-379/Product: core protein p2 #status experimental <PP2>  
F/380-434/Product: core protein p7 #status experimental <PP7>  
F/435-450/Product: core protein p1 #status experimental <PP1>  
F/451-506/Product: core protein p6 #status experimental <PP6>

Alignment Scores:

Pred. No.:	8.71e-103	Length:	506
Score:	2179.50	Matches:	413
Percent Similarity:	88.72%	Conservative:	43
Best Local Similarity:	80.35%	Mismatches:	37
Query Match:	76.74%	Indels:	21
DB:	1	Gaps:	6

US-09-475-704A-4 (1-1509) x A38068 (1-506)

QY 4 GGGCGCCCCGGCACTTCTGCGCGGAGAGCTGACAAGTGGAGAGATCCGCTG 63  
DB 1 GlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTrpGluLysIleArgLeu 20  
QY 64 CGCGCGCGCGGCAAGACACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAGCTG 123  
DB 21 ArgProGlyGlyLysLysLysTyrLysLeuLysHisValValTrpAlaSerArgGluLeu 40  
QY 124 GAGGCTTGGCCCTGAACCCCGGCTGTGAGACCGCGGAGGGCTGCAAGCATCATG 183  
DB 41 GluArgPheAlaIleAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIleLeu 60  
QY 184 AAGCAGCTGCAGCCCGCTGCAGACCGGCAACCGAGAGCTGCGCAGCTGTACAACACC 243  
DB 61 GlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluArgLysSerLeuTyrAnThr 80  
QY 244 GTGGCCACCTGTACTGCTGTGACGCGCGCATCGAGTCCGCGACCAAGAGGCCCTG 303

DB 81 ValAlaThrLeuTyrCysValHisGlnLysIleLysIleLysAspThrLysGluAlaLeu 100  
QY 304 GACAAGATCGAGAGAGAGACAAGTCCACGAGAAAGACCGACGACGAGAGGCC 363  
DB 101 GluLysIleGluGluGluGlnAsnLysSerLysLysAlaGlnGlnAlaAlaAsp 120  
QY 364 GACGGC-----AAGTGAGCCAGAACTACCCCATCGTGACAACTG 405  
DB 121 ThrGlyAsnArgGlyAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIle 140  
QY 406 CAGGCGCAGATGGTGACACGACCATCAGCCCGCACCCCTGAAGCCTGGTGAAGTG 465  
DB 141 GlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysVal 160  
QY 466 ATCGAGAGAGAAGCCTTCAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGC 525  
DB 161 ValGluGluLysAlaPheSerProGluValIlePrometPheSerAlaLeuSerGluGly 180  
QY 526 GCCACCCCCAGACCTGAACACGATGTTGAACACCGTGCGCGCCACGAGCGCCATG 585  
DB 181 AlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMet 200  
QY 586 CAGATGCTGAAGACACCATCAACGAGAGAGCGCCGAGTGGACCGCCTGCACCCCGTG 645  
DB 201 GlnMetLeuLysGluThrIleAsnGluGlnAlaAlaGluTrpAspArgLeuHisProAla 220  
QY 646 CAGGCGCGCGCGCGCGCGCGCGCGAGATGCGCGCACCCCGCGGACGACATCGCGCG 705  
DB 221 HisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGly 240  
QY 706 GCCACCAAGACCTCTGACAGAGACGATCGCTGATGACCAACACCCCGCTGCCCGTG 765  
DB 241 ThrThrSerThrLeuGlnGlnIleGlyTyrMetThrAsnAsnProProIleProVal 260  
QY 766 GGGCAGATCTACAAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGATGACAC 825  
DB 261 GlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSer 280  
QY 826 CCGGTGAGCATCTTGACATCCGCCAAGGCGCCCAAGAGCCCTTCCGCACTACGTGAC 885  
DB 281 ProSerSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAsp 300  
QY 886 CGCTTCTTCAAGACCCCTGCGCGCGGAGCAGGCGCAACCGACGCTGAAGAACTGATGACC 945  
DB 301 ArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThr 320  
QY 946 GAGACCTGTGTGTGACGAACGCGCAACCCGCACTGCAAGACCATCTGCGGCTTCGGC 1005  
DB 321 GluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGly 340  
QY 1006 CCGGCGCGCAACCTTGAGAGAGATGACCGCGCTGCCAGGGCGTGGGGCGCGCGCCAC 1065  
DB 341 ProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHis 360  
QY 1066 AAGCGCGCGTGTGGCGGAGCGATGAGCCAG---GCCAACAACGCTGAACATCATGATG 1122  
DB 361 LysAlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMet 380  
QY 1123 CAGAAAGCAACTTCAAGGGCGCGCGCGCAACGTCAAGTGTCTTCAACTGCGGCAAGAG 1182  
DB 381 GlnArgGlyAsnPheArgAsnGlnArgLysIleIleLysCysPheAsnCysGlyLysGlu 400  
QY 1183 GGGCAGATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGGCTGTGAAGTGGCGCAAG 1242  
DB 401 GlyHisIleAlaLysAsnCysArgAlaProArgLysArgGlyCysTrpLysCysGlyLys 420  
QY 1243 GAGGGCCACCAAGATGAAGACTGCACCGAGCGCGCAAGGCAACTTCTGGGCAAGATCTGG 1302  
DB 421 GluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrp 440  
QY 1303 CCGAGCCACAAGGGCGCGCGCGCAACTTCTTGCAAGAACCGCAGCGAGCGCGCGCGCC 1362

Db 441 ProSerCyblysglyArgProglYAsnPhseProglInserArgThrIu----- 456  
QY 1363 ACCGTGCCACCGCCCCCGCGAGAGCTTCGCTTC-----GAGAGACCACCCCC 1416  
Db 457 -----ProThrAlaProProglIngluSerPheArgPheglYgluThrThrPro 474  
QY 1417 GCCCCCAAGCAGAGCCCAAG-----GACCGCGAGCCCTACCGCGAGCCCTG 1464  
Db 475 TyrGlnlysgIngluLylslysgIngluThrIleAsplyAspLeuTyr-----Proleu 492  
QY 1465 ACCGCGCTGCGCAGCCTGTTCGCGAGCGCCCTGAGCCAG 1506  
Db 493 AlaSerLeuLysSerLeuPheglYAsnAspProLeuSerGln 506

RESULT 12  
FOVWVL

gag polypeptide - human immunodeficiency virus type 1 (isolate LV)  
N/Alternate names: assemblin; core polypeptide; gag precursor  
N/Contains: capsid antigen core protein p24CA; core protein p1; core protein p6; matrix  
C/Species: human immunodeficiency virus type 1, HIV-1  
C/Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004  
C/Accession: A03948  
R/Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.  
Nature 313, 450-458, 1985  
A/Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi  
A/Reference number: A93355; MUID:8511157; PMID:2982104  
A/Accession: A03948  
A/Molecule type: DNA  
A/Residues: 1-478 <MOE>  
A/Cross-references: UNIPROT:P03350  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polypeptide  
C/Keywords: AIDS; blocked amino end; core protein; immunodeficiency; lipoprotein; myrist  
F/2-478/Product: gag precursor (assemblin) #status predicted <GAG>  
F/2-132/Product: matrix antigen core protein p17MA #status predicted <P17>  
F/20-32/Region: nuclear location signal  
F/110-114/Region: nuclear location signal  
F/133-363/Product: capsid antigen core protein p24CA #status predicted <P24>  
F/364-377/Product: core protein p2 #status predicted <CP2>  
F/378-432/Product: nucleocapsid core protein p7NC #status predicted <CP7>  
F/392-405/Region: zinc finger CCHC motif  
F/413-426/Region: zinc finger CCHC motif  
F/2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted  
F/392,395,400,405/Binding site: zinc (Cys, Cys, His, Cys) #status experimental  
F/413,416,421,426/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

Alignment Scores:

Pred. No.: 1.15e-94 Length: 478  
Score: 2018.50 Matches: 382  
Percent Similarity: 87.86% Conservative: 45  
Best Local Similarity: 78.60% Mismatches: 40  
Query Match: 71.07% Indels: 19  
DB: 1 Gaps: 5

US-09-475-704A-4 (1-1509) x FOVWVL (1-478)

QY 1 ATGGGCGCGCGCGCGCATCTCGCGCGCGAGCAAGCTGAGCAAGTGGAAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlygluLeuAspArgTrpGluysIleArg 20  
QY 61 CTGCGCGCGCGCGCGAGCAAGCACTACATGCTGAAGCAACCTGTGTGGCCGCGGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAlaSerArgIu 40  
QY 121 CTGAGAGGCTTCGCTGAAACCGCGCGCTGTGAGAGACCGCGAGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
QY 181 ATGAAGCAGCTGAGCGCGCGCTGAGACCGCGAGCTGCGAGCTGTACAAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluLeuArgSerLeuTyrAsn 80

QY 241 ACCGTGCCACCGCTGTACTGCGTGACGCGCGCATCGAGGTCCGCGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100  
QY 301 CTGACAAGATTCGAGGAGAGGACGACAACTCCACGAGAGACCAGAGCGCAAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGlnAlaAla 120  
QY 361 GCCGACGCG-----AAGGTGAGCCAGAACTAACCCCATGCTGCAAGACCTGCAAGGC 411  
Db 121 AspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly 140  
QY 412 CAGATGTGCACGAGGCCATCAGCCCCCGCACCTGAAACCGCTGGGTGAAGTGATCGAG 471  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleTrpValLysValGlu 160  
QY 472 GAGAGGCTTCAGCGCGCGAGTGATCCCATGTTCACCGCCCTGAGCGAGCGCCACC 531  
Db 161 GluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThr 180  
QY 532 CCCGAGACCTGAAACAGATGTGAACACCGTGGCGCGCCACGAGCGCCCATGCAATG 591  
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMet 200  
QY 592 CTGAAGGACCATCAACGAGAGAGCGCGCGAGTGAGACCGCCTGCAACCGCTGCAAGGC 651  
Db 201 LeuLysgluThrIleAsnGluGlnAlaIleGluTrpAspArgValHisProValHisAla 220  
QY 652 GCGCGCGTGGCGCGCGCGCATGCGCGACCGCGCGCGCGCGCATCGCGCGCGCCACC 711  
Db 221 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAglYThrThr 240  
QY 712 AGCACCTGCGAGAGCAGATCGCTGATGACCAACCGCCCGCTGCGCGCGAG 771  
Db 241 SerThrLeuGlnGluGlnIleGlyTyrMetThrAsnAspProIleProValGlyGlu 260  
QY 772 ATCTACAAGCGGTGATCATCTCGCGCTGAACAAGATCGTGGATGATACGCGCGGTG 831  
Db 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProthr 280  
QY 832 AGCATCTGACATCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 891  
Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAAGACCTGCGCGCGCGAGCAGCGCACCGCAGAGCTGAAAGACTGATGACCGAGACC 951  
Db 301 TyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGluThr 320  
QY 952 CTGCTGTGCGAGAGCGCAACCGCGCATGCAAGCACTCTGCGCGCTCTCGCGCGCGC 1011  
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340  
QY 1012 GCCACCTGAGAGAGATGATGACCGCGCTGCCAGGCGGTGGGGCGCGCGCGCGCAAGGCC 1071  
Db 341 AlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla 360  
QY 1072 CGCGTGTGCGCGAGCGCGATGAGCCAG---GCCAAGCGGTGAACATCATGATGAGAG 1128  
Db 361 ArgValLeuAlaGluAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGlnArg 380  
QY 1129 AGCAACTTCAAGGCGCGCGCGCGCAAGCTCAAGTGTCACTGCGCGCAAGAGGCGCAC 1188  
Db 381 GlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnCyseGlyLysGluGlyHis 400  
QY 1189 ATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGC 1248  
Db 401 ThrAlaArgAsnCyseArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGly 420  
QY 1249 CACCAAGTGAAGACTGACCGAGCGCGCGCAACTTCTGGGCAAGATC---TGCGCC 1305  
Db 421 HisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleCysLeuPro 440  
QY 1306 AGCCACAAGGCGCGCGCGCAACTTCTGACAGAACCGGAGAGCGCGCGCGCGCCACC 1365

```

Db      441 ThrArggluGlynglnGlyIlePhePheArgAlaAspGlnSerGln----- 455
QY      1366 GTGCCCAACGGCCCCCCCCCGCCAGAGCTTCGGCTTCGAGAGACCACCCCCGCCCAAG 1425
Db      456 -----GlnProHisHisPhePheArgAlaAspGlnSer----- 466
QY      1426 CAGGAGCCCAAGGACCGC 1443
Db      467 GlnGlnProHisGlnIlys 472

```

## RESULT 13

gag polyprotein - simian immunodeficiency virus SIVcpz  
N:Alternate names: core polyprotein  
C:Species: simian immunodeficiency virus SIVcpz  
A>Note: host Pan troglodytes (chimpanzee)  
C>Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C/Accession: S09983  
R:Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.  
Nature 345, 356-359, 1990  
A>Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.  
A:Reference number: S09983; MUID:90259077; PMID:2188136  
A:Accession: S09983  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-508 <HUE>  
A:Cross-references: UNIPROT:P17282; EMBL:X52154; NID:g58866; PIDN:CAA36401.1; PID:g58867  
C:Genetics:  
A:Gene: gag  
C:Superfamily: AIDS-related virus gag polyprotein  
C:Keywords: AIDS; core protein; immunodeficiency; polyprotein

### Alignment Scores:

Pred. No.:	7.19e-92	Length:	508
Score:	1963.00	Matches:	357
Percent Similarity:	82.56%	Conservative:	69
Best Local Similarity:	69.19%	Mismatches:	68
Query Match:	69.12%	Indels:	22
DB:	1	Gaps:	4

US-09-475-704A-4 (1-1509) x FOLJSI (1-508)

[illegible]

Db	141	ProLeuValGlnAsnAlaGlnGlnGlnMetValHisGlnAlaIleSerProArgThrLeu	160
OY	448	AACGCTGGGTGAAGGTGATCGAGAGAAAGCCTTCAGCCCCGAGGTGATCCCATGTTTC	507
Db	161	AsnAlaTrpValIlySValValGlnGlnIlySAlaPheSerProGluValIleProMetPhe	180
OY	508	ACCGCCCTGAGCGAGGGCGGCCACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGC	567
Db	181	SerAlaLeuSerGlnGlnIlyAlaLeuProGlnAsnPrpValAsnThrMetLeuAsnAlaValGly	200
OY	568	GGCCACCAAGGCGCCATGCGAGATGCTGAAGGACACCATCAACGAGGAGGCGCCGAGTGG	627
Db	201	GlyHisGlnGlnIlyAlaMetGlnValLeuIlySglnValIleAsnGlnGlnAlaIlaGlnTrp	220
OY	628	GACCGCTGCAACCCCGTGCAAGGCCGGCCCCCGTGGCCCCCGCCAGATGCCGACCCCCGC	687
Db	221	AspArgLeuHisProThrHisAlaGlnProIleAlaProGlnGlnLeuArgGluProArg	240
OY	688	GCGAGCGACATGCGCGCGGCCACCGACACCTGCAGAGAGAGATGCGCTGATGACCAAC	747
Db	241	GlySerAspIleAlaGlnThrThrSerThrLeuGlnGlnIleGlyTrpThrThraIa	260
OY	748	AACCCCCCGTGGCCGTTGGGCGCATCTCAAGACGGGTGATCATCTGGGCTGAACAAG	807
Db	261	AsnProProIleProValGlnAspValIlyTrpArgArgTrpValIleLeuGlnLeuAsnIlyS	280
OY	808	ATCGTGGGATGTACAGCCCCGTGAGCATCTGACATCCGCCAGGGCCCCCAAGAGCCC	867
Db	281	ValValArgMetIlyCysProValIserIleLeuAspIleArgGlnIlyProIlySglnPro	300
OY	868	TTCCGCGACTACGTGACCGCTTCTTCAAGACCTGTGCGCGCGAGCAGAGGCCACCCAGGAC	927
Db	301	PheArgAspIlyTrpValAspArgPheIlyThrLeuArgAlaGlnGlnAlaSerGlnGln	320
OY	928	GTGAAGAAGTGGATGACCGCAGACCCCTGCTGTGTCAGAACGCCAACCCCGACTGCAAGACC	987
Db	321	ValIlySAsnTrpMetThrAspThrLeuLeuValGlnAsnAlaAsnProAspCylSylSgln	340
OY	988	ATCCTGCGCGCTCTCGGCCCCCGCGCCACCTGAGAGAGATGATGACCGCTGCCAGGGC	1047
Db	341	IleLeuIlySAlaLeuGlnIlyProGlnIlyAlaThrLeuGlnGlnMetThrAlaCylSglnGly	360
OY	1048	GTEGGCGGCCCCCGCCACAAGGCCCGCGTGTGGCCGAGCGATGAGCCAGCCCAACAGC	1107
Db	361	ValGlyGlyProSerHisIlySAlaArgValLeuAlaGlnAlaMetSerMetValGlnAsn	380
OY	1108	-----GTGAACATCATGATGCAGAAAGCAACTTCAAGGGCCCCCGCGCCAAAGTC	1158
Db	381	GlnGlyArgAlaAspValPhePheGlnIlySglnGlnIlyAlaGlnProIlyAspArgIlySile	400
OY	1159	AAGTGTTCAACTGCGGCAAGGAGGGCCACATGCCCAAGAACTGCCGCGCCCCCGCAAG	1218
Db	401	IlySAspPheAsnCylSylIlySglnIlyHisLeuAlaArgAsnCylSylAlaProArgArg	420
OY	1219	AAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCATGAAGACTGCACCCGAGCGCCAG	1278
Db	421	IlySglnCylTrpArgCylSylGlnGlnIlyHisGlnMetIlySAspCylThrGlnIlyArgGln	440
OY	1279	GCCAACTTCTGGGCAAGATCTGGCCCAAGCCAAAGGGCGCCCCCGCAACTTCTGCAAG	1338
Db	441	ValAsnPheLeuGlnIlySglnIlyTrpProSerArgSerGlnIlyArgProGlnAsnPheValGln	460
OY	1339	AACCGCAGCGAGCCCGCGCCCCCAACCGTGCACCGCCCCCCCCCGCCAGAGCTTCCGC	1398
Db	461	AsnArgThrGln-----ProThrAlaProProIleGlnIlySerIlyGly	474
OY	1399	TTCGAGGAGAACACCCCGCGCCCCCAAGCAGAGAGCCCAAGAGCCGAGCCCTACCGCGAG	1458
Db	475	TyrGlnGlnGlnIlyIlySAsp-----GlnGlnIlySylSglnGlnIlyGlnIlySerLeuTyr	492
OY	1459	CCCCTGACCGCCCTGCGCAGCTGTTGGGAGCGGGCCCCCTGAGCCAG	1506
Db	493	ProProThrSerLeuIlySAspThrLeuPheGlnIlySerAspProSerSerGln	508

RESULT 14  
S46346  
gag polypeptide - simian immunodeficiency virus SIVagm (isolate SAB-1)  
C:Species: simian immunodeficiency virus SIVagm  
A:Variety: isolate SAB-1  
C:Date: 25-Dec-1994 #sequence\_revision 14-Feb-1997 #text\_change 26-Aug-1999  
C:Accession: S46346  
R:Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.;  
EMBO J. 13, 2935-2947, 1994  
A:Title: Mosaic genome structure of simian immunodeficiency virus from West African green  
A:Reference number: S46335; MUID:94298785; PMID:8026477  
A:Accession: S46346  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-554 <JIN>  
A:Cross-references: EMBL:U04005; NID:9466229; PIDN:AAA21504.1; PID:9466230  
A:Experimental source: isolate SAB-1; babaeus monkey  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993  
C:Genetics:  
A:Gene: gag  
C:Superfamily: AIDS-related virus gag polypeptide  
C:Keywords: polypeptide

Alignment Scores:  
Pred. No.: 1.17e-63 Length: 554  
Score: 1403.50 Matches: 290  
Percent Similarity: 65.18% Conservative: 75  
Best Local Similarity: 51.79% Mismatches: 122  
Query Match: 49.42% Indels: 73  
DB: 2 Gaps: 14

US-09-475-704a-4 (1-1509) x S46346 (1-554)

```
QY 1 ATGGGCGCCCGCGGCATCTGCGCGCGAGAGCTGACAAAGTGGAGAAATCCGC 60
DB 1 MetGlyAlaSerAsnSerValLeuSerGlyArgLysLeuAspAlaPheGluSerValArg 20
QY 61 CTGGCGCCCGCGCGAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCCGCGAG 120
DB 21 LeuArgProAsnGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 121 CTGAGAGGCTTCCGCTGAACCCCGGCTGTGAGAGCCGCGAGGCTGCAAGCATC 180
DB 41 LeuAspArgPheSerLeuSerAlaAsnLeuGluThrLysGluGlyValValLysLys 60
QY 181 ATGAAGCAGCTGACAGCCCGCTGAGAGCCGCGAGAGGCTGCGCAGCTGTACAAC 240
DB 61 LeuSerValLeuLeuProLeuValProThrGlySerGluAsnLeuIleAlaLeuPheAsn 80
QY 241 ACCGTGGCCACCTCTGTACTGCGTGCAGCGCGCATGAGGTCCGCGACACCAAGAGGCC 300
DB 81 LeuCySerValLeuAlaCysIleHisAlaGluIleLysValLysAspThrGluGluAla 100
QY 301 CTGACAAAGATCGAGAGAGG-----CAGAAC 327
DB 101 LysAlaLysValLysGluGluValProAlaGluMetThrGluSerAlaThrAlaThrSer 120
QY 328 AAGTCCAGCAGAAAGACCCAGCAGGCCAAGAGGCCGAGCGCAAGGTG----- 375
DB 121 SerGlyGlnThrLysGluLeuGlnAlaLysLysLysAsnGluProThrValThrProSer 140
QY 376 -----AGCCAGAACTACCCCATGTCAGAACTGCAAGGCGCAGATGTGCAAGGCC 429
DB 141 GlyLysSerArgAsnTyrProIleVal--SerValAsnAsnGlnTrpValHisGlnPro 159
QY 430 ATCAGCCCCCGGACCTGAAGCGCTGGGTGAAGGTGATCGAGGAGAAAGGCTTCAGCCCC 489
DB 160 LeuSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysLysPheSerAla 179
QY 490 GAGGTGATCCCATGTTACCGCGCTGAGCGAGGGCGCCACCCCGCAGACCTGAACACG 549
DB 180 GluValValProMetPheSerAlaLeuAlaGluGlyAlaIleProTyrAspIleAsnGln 199
```

```
QY 550 ATGTTGAACACCGTGGCGCGCCACCGCCCATGACATGCTGAAGACACCATCAAC 609
DB 200 MetLeuAsnAlaValGlyGluHisGlnGlyAlaLeuGlnIleValLysAspValIleAsn 219
QY 610 GAGAGAGCCCGGAGTGGAGCCGCTGCACCCCGGTGACGGCGCCCGGCTGCGCCCGC 669
DB 220 GluGluAlaAlaAspTrpAspLeuArgHisProProGlnGlnProProAlaGlnGly 239
QY 670 CAGATCGCGACCCCGCGGAGCGACATGCGCGCGCCAGCAGACCCCTGAGAGAGCAG 729
DB 240 ValLeuArgAspProGlnGlySerAspIleAlaGlyThrThrSerThrIleGlnGluGln 259
QY 730 ATGCGCTGATGACC--AGCAACCCCGCGTGGCGCGACATCTACAGCGGTGG 786
DB 260 IleGluTrpThrThrArgAlaGlnAsnAlaValAsnValGlyAsnIleTyrLysGlyTyr 279
QY 787 ATCATCTGGGCTGAACAAAGATCGTGCAGATGTACAGCCCGGTGACATCTGACATC 846
DB 280 IleIleLeuGlyLeuGlnLysCysValLysMetTyrAsnProValAsnIleLeuAspIle 299
QY 847 CGCCAGGGCCCGCAAGAGACCTTCGCGACTACGTGACCGCTTCTTCAAGACCTGCGC 906
DB 300 LysGlnGlyProLysGluProPheLysAspTyrValAspArgPheTyrLysAlaLeuArg 319
QY 907 GCCGAGCAGGCCACCCAGAGCGTGAAGAACTGATGACCGAGACCCCTGTGTCAGAAC 966
DB 320 AlaGlnGlnThrAspProAlaValLysAsnTrpMetThrGlnSerLeuLeuIleGlnAsn 339
QY 967 GCCAACCCCGCTGCAAGACCATCTGCGCGCTCTCGCGCCCGCGCCACCTGAGAGAG 1026
DB 340 AlaAsnProAspCysLysThrValLeuLysGlyLysMetAsnProThrLeuGluGlu 359
QY 1027 ATGATGACCGCTGCCAGGCGGTGGCGCGCCCGCCACAGAGCCCGCTGTGCGCGAG 1086
DB 360 MetLeuThrAlaCysGlnGlyIleGlyGlyAlaGlnHisLysAlaArgLeuMetAlaGlu 379
QY 1087 GCGATGAC-----CAGGCCAAGCGGTGAACATCATGATGAGAG--AGC 1131
DB 380 AlaMetThrAlaAlaPheGlnGlnGlnThrValGlyAsnIlePheValGlnGlnGlyAla 399
QY 1132 AACTCAAGGGCCCC-----CGCGCAACGTCAGAGTGCCTTC 1167
DB 400 ArgProArgGlyProLeuGlnGlyArgGlyArgProLeuAsnProAsnIleLysCysTyr 419
QY 1168 AACTGCGGCAAGAGGGGCCATCGCCCAAGACTGCGCGCCCGCCGCAAGAGGGCTGC 1227
DB 420 AsnCySerGlyLysProGlyHisLeuAlaArgPheCysLysAlaProArgArgGlnGlyCys 439
QY 1228 TGAAGTGGCGCAAGAGGGGCCACCAAGTGAAGACTGCAACCGCGCGCAAGGCAACTTC 1287
DB 440 TrpLysCysGlySerProAspHisGlnMetLysAspCys--GlnLysGlnValAsnPhe 458
QY 1288 CTGGGCAAGATCTGGCCAGCCACCAAGAGGGCGCCCGGCAACTTCTGCAAGACCGCAGC 1347
DB 459 LeuGly---PheGlyProTrpGlyArgGlyLysProArgAsnPhe----- 472
QY 1348 GAGCCCGCGCCCGCCACCGTGCACCGCCCGCCCGCGGAGAGC-----TTCGCTTC 1401
DB 473 ---ProLeuThrSerIleArgProThrAlaProProMetGluArgAspTyrSerArgPro 491
QY 1402 GAGGAGACC-----ACCCCGCGCCCG----- 1422
DB 492 GluGluAsnTrpTyrAlaAspArgProProThrArgGlyProGlyProAspAspProAla 511
QY 1423 -----AAGCAGAG----- 1431
DB 512 ThrAlaLeuLeuLysGlnTyrAlaValGlnGlyLysArgGlnLysGlnGlnTrpGlnAsn 531
QY 1432 ---CCCAAGAGACCGCGACCTACCGCGAGACCCCTGAGCGACCGCTGTGCGGC 1488
DB 532 HisSerProGlnGlnSerProTyrGlnGlnAlaTyrSerSerLeuArgSerLeuPheGly 551
```

```
RESULT 15
FOLJTM
gag polyprotein - simian immunodeficiency virus (strain stm)
N/Alternate names: core polyprotein
N/Contains: core protein p17; core protein p24
C/Species: simian immunodeficiency virus, Siv
A/Note: host Macaca arctoides (stump-tailed macaque)
C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C/Accession: A41565
R/Khan, A.S.; Galvin, T.A.; Lowenstine, L.J.; Jennings, M.B.; Gardner, M.B.; Buckler, C.
J. Virol. 65, 7061-7065, 1991
A/Title: A highly divergent simian immunodeficiency virus (SIVstm) recovered from stored
A/Reference number: A41565; M0ID:92046379; PMID:1942258
A/Accession: A41565
A/Molecule type: DNA
A/Residues: 1-510 <KHA>
A/Cross-references: UNIPROT:P31634; GB:X60667; NID:g60527; PIDN:CAA43084.1; PID:g60528
C/Genetics:
A/Gene: gag
C/Superfamily: AIDS-related virus gag polyprotein
C/Keywords: core protein; polyprotein
F/1-115/Product: core protein p17 #status predicted <P17>
F/116-510/Product: core protein p24 #status predicted <P24>

Alignment Scores:
Pred. No.: 2.81e-63 Length: 510
Score: 1396.00 Matches: 279
Percent Similarity: 67.88% Conservative: 76
Best Local Similarity: 53.35% Mismatches: 128
Query Match: 49.15% Indels: 40
DB: 1 Gaps: 11

US-09-475-704A-4 (1-1509) x FOLJTM (1-510)
QY 1 ATGGGCGCCGCGCGCATCTCGCGCGGAGAGCTGACCAAGTGGAGAGAAGATCCGC 60
Db 1 MetGlyAlaArgSerSerValLeuSerGlyLysLysAlaAspGluLeuGluLysValArg 20
QY 61 CTGGCGCCCGCGCGGAGAGCACTACGTGTAAGCACTGTGGAGCCAGCCGCGAG 120
Db 21 LeuArgProGlyLysLysLysLysLysMetLeuLysHisValValTrpAlaAlaAsnGlu 40
QY 121 CTGAGGGGCTTCCGCTGAACCCCGGCTGTGAGACCGCGAGGCTGCAAGCAGATC 180
Db 41 LeuAspArgPheGlyLeuAlaGluSerLeuGluSerLysGluGlyCysGlnLysIle 60
QY 181 ATGAAGCAGCTGACGCGCGCGCTGACAGCGGACGAGGAGCTGGCGAGCTGTACAAC 240
Db 61 LeuThrValLeuGluProLeuValProThrGlySerGluAsnLeuLysSerLeuPheAsn 80
QY 241 ACCGTGGCCACCCCTACTGCGTGCAAGCGCGGATGAGGTCCGCGACACCAAGAGGCC 300
Db 81 ThrValCysValIleTrpCysIleHisAlaGluGluValLysHisThrGluGluAla 100
QY 301 CTGACACAAGATCGAGGAGAG-----CAGAACAACTCCACGACAG 339
Db 101 LysGlnValValLysArgHisLeuValValGluThrGlyThrAlaAspLysMetProAla 120
QY 340 AAGACCCAGCAGGCGCAAGAGGCGCGGCAAGGTGACCAAGACTACCCCATCGTCAG 399
Db 121 ThrSerArgProThrAlaProProSerGlyArg---GlyGlyAsnTyrPro---ValGln 138
QY 400 AACTGCAAGGCGCAAGTGTGCAACCAAGCCATCAGCCCGGACCCCTGAAGCGCTGGTG 459
Db 139 GlnValGlyGlyAsnTyrValHisLeuProLeuSerProArgThrLeuAsnAlaTrpVal 158
QY 460 AAGTGATCGAGGAGAAGGCTTCAGCCCGAGGTGATCCCATGTTCAACGCGCTGAGC 519
Db 159 LysLeuValGluGluLysLysPheGlyAlaGluValValSerGlyPheGlnAlaLeuSer 178
QY 520 GAGGCGCCACCCCGGACGCTGAACACGATGTTGAACACCGTGGCGGCGCACCGAGCC 579
Db 179 GluGlyCysThrProTyrAspIleAsnGlnMetLeuAsnCysValGlyLysIleGlnAla 198
```

```
QY 580 GCCATGCAGATGCTGAAGACACCACTCAACGAGAGCGCGCGAGTGGAGCCGCTGCAC 639
Db 199 AlaMetGlnIleIleArgGluIleIleAsnGluAlaAlaAspTrpAspValGlnHis 218
QY 640 CCCGTGAGGCGCGCGCGTGGCGCGCGGACAGATGCGCGAGCCCGCGGACGAGCATC 699
Db 219 Pro---GlnProGlyProLeuProAlaGlyGlnLeuArgGluProSerGlySerAspIle 237
QY 700 GCGGCGCCACGACCAACCTGACAGAGCAGATCGCTGATG---ACCAGCAACCCCC 756
Db 238 AlaGlyThrThrSerThrValGluGluGlnIleGlnTrpMetHisArgGlnGlnAsnPro 257
QY 757 GTGCCCGTGGCGGACATCTACAAGCGGTGATCATCTGGGCTGGAACAAGATCGTGGG 816
Db 258 IleProValGlyAsnIleTyrArgArgTrpIleGlnLeuGlyLeuGlnLysCysValArg 277
QY 817 ATGTACAGCCCGGTGAGCATCTCGGACATCCGCGAGGCGCCCAAGAGCCCTTCGCGAC 876
Db 278 MetTyrAsnProValAsnIleLeuAspIleLysGlnGlyProLysGluProPheGlnSer 297
QY 877 TACGTGACCGCGCTTCTTCAAGACCTGCGCGCGCGGACGACCCACCAAGACGTGAAGAC 936
Db 298 TyrValAspArgPheTyrLysSerLeuArgAlaGluGlnAlaAspProAlaValLysAsn 317
QY 937 TGGATGACCGAGACCCCTGTGTGAGAGACGCCCAACCCCGACTGCAAGACCATCTGCGC 996
Db 318 TrpMetThrGlnThrProLeuIleGlnAsnAlaAsnProAspCysLysLeuValLeuLys 337
QY 997 GCTCTCGCGCGCGCGCGCACCCCTGAGAGAGATGATGACCGCTGCCAGGGCGTGGCGGC 1056
Db 338 GlyLeuGlyMetAsnProThrLeuGluGluMetLeuThrAlaCysGlnGlnValGlyGly 357
QY 1057 CCGGCGCCACAAAGCGCGCGTGTGTGAGCGGAGCGGATGAGCCAGCCCAAGCGTGAATC 1116
Db 358 ProGlyGlnLysAlaArgLeuMetAlaGluAlaLeuLysGluAlaPheGlnProGlyPro 377
QY 1117 ATGATGCAAGAGCAACTTCAAGGCGCGCGCGGCGCAACGTCAAGTGTCTCACTGCGGC 1176
Db 378 LeuProPheAlaAlaAlaGlnGlnGlnGlyArgArgThrValLysCysTrpAsnLysGly 397
QY 1177 AAGAGGGCCACATCGCCCAAGACTGCGCGCGCGCGCGGACGAGAGGCTGTGGAATGC 1236
Db 398 LysGluGlyHisThrAlaLysGlnCysLysAlaProArgArgGlnGlyCysTrpLysCys 417
QY 1237 GGCAGGAGGCGCCACCAAGATGAAGAACTGACCGGAGCGCGGCAACTCTCTGGGCAAG 1296
Db 418 GlyLysProGlyHisGlnMetAlaLysCysProGluArgGlnValGlyPheLeuGly--- 436
QY 1297 ATCTGGCCCGACCAAGGCGCGCGCGCGGCAACTC-----CTGCAGAAC 1341
Db 437 PheGlyProTrpGlyLys---LysProArgAsnPheProMetAlaGlnIleProGlnGly 455
QY 1342 CGGAGCGAGCGCGCGCGCGCGCGGCTGCGGAGCGCGCGCGCGGAGAGCTTCGCTTC 1401
Db 456 LeuThrProThrAlaProProGluMetProThrAlaProProValAsp----- 471
QY 1402 GAGGAGACCAACCCCGC-----CCCAAGCAG 1428
Db 472 -----ProAlaAlaAspLeuLeuArgSerTyrMetGlnLeuGlyLysLysGln 487
QY 1429 GAGCCCAAGACCGCGAGCGCTTACCGC-----GAGCCCTGACCGCGCTGGCAGC 1479
Db 488 ArgGluSerArgLysThrProTyrLysGluValThrGluAspLeuValHisLeuAsnSer 507
QY 1480 CTGTTGCGC 1488
Db 508 LeuPheGly 510
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:39:57 ; Search time 174.728 Seconds  
(without alignments)  
5584.002 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 2772  
Sequence: 1 atcgggcgccgcgcgcagcat.....acgacccctgagccagctaa 1479

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1396920 seqs, 329844858 residues

Total number of hits satisfying chosen parameters: 2793840

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool\_p/US09475704/runat\_10032005\_140224\_149377/app\_query.fasta\_1.3342  
-DB=Published\_Applications\_AA -QFMT=faстан -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09475704\_@CCN\_1\_466\_@runat\_10032005\_140224\_14937  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result \*  
Query

No.	Score	Match	Length	DB	ID	Description
1	2606	94.0	492	10	US-09-899-575-17	Sequence 17, Appl
2	2448	88.3	492	16	US-10-332-413-2	Sequence 2, Appl
3	2435	87.8	492	9	US-09-991-258-5	Sequence 5, Appl
4	2435	87.8	631	16	US-10-332-413-8	Sequence 8, Appl
5	2423	87.4	492	14	US-10-339-217-110	Sequence 110, App
6	2416	87.2	492	14	US-10-339-217-109	Sequence 109, App
7	2415.5	87.1	491	14	US-10-339-217-140	Sequence 140, App
8	2407.5	86.9	491	14	US-10-339-217-104	Sequence 104, App
9	2396.5	86.5	491	14	US-10-339-217-143	Sequence 143, App
10	2382	85.9	496	14	US-10-339-217-108	Sequence 108, App
11	2381.5	85.9	487	14	US-10-339-217-106	Sequence 106, App
12	2381.5	85.9	497	14	US-10-339-217-107	Sequence 107, App
13	2344	84.6	508	14	US-10-339-217-105	Sequence 105, App
14	2343	84.5	502	10	US-09-899-575-22	Sequence 22, Appl
15	2328.5	84.0	494	14	US-10-339-217-126	Sequence 126, App
16	2297	82.9	500	14	US-10-339-217-18	Sequence 18, Appl
17	2277	82.1	498	16	US-10-325-468-22	Sequence 22, Appl
18	2266	81.7	500	16	US-10-325-468-8	Sequence 8, Appl
19	2263	81.6	500	16	US-10-325-468-34	Sequence 34, Appl
20	2262	81.6	498	16	US-10-325-468-27	Sequence 27, Appl
21	2258	81.5	500	14	US-10-339-217-121	Sequence 121, App
22	2256	81.4	500	14	US-10-097-534-32	Sequence 32, Appl
23	2256	81.4	500	14	US-10-059-271-85	Sequence 85, Appl
24	2256	81.4	500	14	US-10-102-622-4	Sequence 4, Appl
25	2256	81.4	500	17	US-10-844-658-6	Sequence 6, Appl
26	2252.5	81.3	499	15	US-10-296-734-1	Sequence 1, Appl
27	2252	81.2	500	16	US-10-325-468-17	Sequence 17, Appl
28	2251	81.2	500	9	US-09-968-355-26	Sequence 26, Appl
29	2251	81.2	583	9	US-09-968-355-17	Sequence 17, Appl
30	2249	81.1	500	14	US-10-223-172A-42	Sequence 42, Appl
31	2249	81.1	500	14	US-10-224-999A-3483	Sequence 3483, Ap
32	2246.5	81.0	501	14	US-10-339-217-136	Sequence 136, App
33	2245	81.0	1350	10	US-09-952-060-35	Sequence 35, Appl
34	2245	81.0	1350	16	US-10-380-641-35	Sequence 35, Appl
35	2237	80.7	557	14	US-10-339-217-144	Sequence 144, App
36	2233	80.6	500	14	US-10-339-217-122	Sequence 122, App
37	2226.5	80.3	503	14	US-10-339-217-127	Sequence 127, App
38	2225	80.3	493	14	US-10-339-217-111	Sequence 111, App
39	2221.5	80.1	499	17	US-10-844-658-2	Sequence 2, Appl
40	2213.5	79.9	494	14	US-10-339-217-100	Sequence 100, App
41	2212.5	79.8	513	14	US-10-339-217-31	Sequence 31, Appl
42	2211.5	79.8	494	14	US-10-339-217-99	Sequence 99, Appl
43	2204.5	79.5	494	14	US-10-339-217-101	Sequence 101, App
44	2199.5	79.3	496	14	US-10-339-217-32	Sequence 32, Appl
45	2195	79.2	512	16	US-10-093-953A-25	Sequence 25, Appl

#### ALIGNMENTS

US-09-899-575-17  
; Sequence 17, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megele, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Estrelita Janse  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; PRIOR FILING DATE: 2001-07-05  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus

US-09-899-575-17

## Alignment Scores:

Alignment Scores:	2,666-129	Length:	492
Pred. No.:	2606.00	Matches:	492
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	94.01%	Gaps:	0
DB:	10		

US-09-475-704a-3 (1-1479) x US-09-899-575-17 (1-492)

```
QY 1 ATGGGCGCGCGCCAGCATCTGCGCGCGCAAGCTGGAAGCCTGGAGCGCATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspAlaTrpGluArgIleArg 20
QY 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACCTGGTGTGGCGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysCysTyrMetMetLysIleLeuValTrpAlaSerArgGlu 40
QY 121 CTGGAAGAAGTTCCCTGAAACCCCGCTGTGAGAGCAGCGAGGCTGCAAGCAGATC 180
Db 41 LeuGluLysPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
QY 181 ATCCGCCAGCTGACACCCCGCTGCAGACCGCGAGCGAGAGCTGAAGAGCTGTTCAC 240
Db 61 IleArgGlnLeuHisProAlaLeuGlnThrGlySerGluGluLeuLysSerLeuPheAsn 80
QY 241 ACCGTGGCCACCTGTACTGCTGTGCACAGAGAAGATCGAGTCCGCGACCAAGAGGCG 300
Db 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGlnAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTGCAGAGCAAGATCCAGAGGCGCGAGCC 360
Db 101 LeuAspLysIleGluGluGlnGlnAsnLysCysGlnGlnLysIleGlnGlnAlaGlnAla 120
QY 361 GCCGACAAAGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAAGCCTGCAAGGCGCAGATG 420
Db 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 140
QY 421 GTGCACCAAGGCCATCAGCGCCCGCACCTGAAAGCCTGGGTGAAGGTGATCGAGAGAAG 480
Db 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLys 160
QY 481 GCCTTCAGCGCCGAGGTGATCCCATGTGTCAACCGCTGAGCGAGGCGCCACCCCGCAG 540
Db 161 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 180
QY 541 GACCTGAACAAGATGTGAACAACCGTGGGCGCGCACCAAGCGCCCATGACATGCTGAAG 600
Db 181 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 200
QY 601 GACACCATCAACGAGAGAGGCGCGCGAGTGGGAGCCGCTGCACCCCGCTGCAGCGCGGCC 660
Db 201 AspThrIleAsnGluGluAlaIleGluTrpAspArgValHisProValHisAlaGlyPro 220
QY 661 ATCGCCCCCGCGCAGATGCGCGAGCGCGCGCGAGCGAGCATCGCGCGCACCAAGCAGACC 720
Db 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240
QY 721 CTGCAAGAGCAGATCGCGCTGGATGACCAACCCCGCATCCCGCTGGGCGCAGATCTAC 780
Db 241 LeuGlnGluGlnIleAlaIleTrpMetThrSerAsnProIleProValGlyAspIleTyr 260
QY 781 AAGCGGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAAGATC 840
Db 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 280
QY 841 CTGGAATCAAGCAGGCGCCCAAGAGGCGCTTCGCGGACTACGTGAAGCGCTTCTCAAG 900
Db 281 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPheLys 300
QY 901 ACCCTGCGCGCGCGAGCAGACCAAGAGGAGTGAAGAACTGATGACCGACACCTGCTG 960
```

```
Db 301 ThrLeuArgAlaGluGlnSerThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 320
QY 961 GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTGTGGCGCGCGCAGC 1020
Db 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340
QY 1021 CTGGAAGAGATGATGACCGCTGCGCAGGCGCGTGGGCGGCGCCAGCACAAGGCGCGCTG 1080
Db 341 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 360
QY 1081 CTGCGCGAGCGGATGAGCCAGGCGCAACCAACGCTGATGATGACAGAGCAACTTCAAG 1140
Db 361 LeuAlaGluAlaMetSerGlnAlaAsnThrSerValMetMetGlnLysSerAsnPheLys 380
QY 1141 GCGCGCGCGCGCATGCTCAAGTCTTCAACTGCGGCAAGAGGCGCACATGCGCGCAAC 1200
Db 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 400
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGCTGTGAAAGTGCAGAGCGCAAGAGGCGCAACAGTGAAG 1260
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleGlnMetLys 420
QY 1261 GACTGCAACCGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGGCGCAAGCAGGCGCGCGC 1320
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGCGCAACTTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1380
Db 441 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProAlaGluSerPheArg 460
QY 1381 TTCAGAGAGACCAACCCCGCGCGCAAGAGCAGAGAGCAGAGCGCGAGACCTGACAGC 1440
Db 461 PheGluGluThrThrProGlyGlnLysGlnGlnSerLysAspArgGluThrLeuThrSer 480
QY 1441 CTGAAGAGCGCTGTTCGCAACGAGCCCGCTGAGCGCAG 1476
Db 481 LeuLysSerLeuPheGlyAsnAspProLeuSerGln 492
```

## RESULT 2

US-10-332-413-2

; Sequence 2, Application US/10332413  
; Publication No. US2004011660A1

## GENERAL INFORMATION:

; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swanstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdool  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
; FILE REFERENCE: 45669-28193  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; PRIOR FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
; US-10-332-413-2

## Alignment Scores:

Alignment Scores:	5,5e-121	Length:	492
Pred. No.:	2448.00	Matches:	461
Score:			



DB: 9 Gaps: 2  
US-09-475-704A-3 (1-1479) x US-09-991-258-5 (1-492)

QY 1 ATGGGCGCGCGCCAGCATCTGCGCGCGCAAGCTGACGCCCTGGAGCGCATCCGC 60  
Db 1 MetAlaAlaArgAlaSerIleLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20

QY 61 CTGGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACCTGTGTGGCGCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisTrpMetLeuLysHisIleValTrpAlaSerArgGlu 40

QY 121 CTGGAAGAATTGCGCCCTGAACCGCGCTGTGAGAGCCAGCGAGGCGTGAAGCATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60

QY 181 ATCCGCGAGCTGCACCCCGCGCTGCAGACCGCGCAGAGAGCTGAAGAGCCTGTTCAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTrpAsn 80

QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAGATCGAGGTCCGCGACCAAGAGGCGC 300  
Db 81 ThrValAlaThrLeuTrpCysValHisGluLysIleGluValArgAspThrLysGluAla 100

QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTGCAGCAGAGATCCAGAGCGCGAGCGC 360  
Db 101 LeuAspLysIleGluGluGlnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120

QY 361 GCGGACAAGGCGAAGGTGAGCCAGAACTACCCATCGTGCAGAACTTCAGAGGCGCAGATG 420  
Db 121 AlaAsp---GlyLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyGlnMet 139

QY 421 GTGACCAAGGCCATCAGCGCGCGCAGCCCTGAACGCGCTGAGAGGTGATCGAGAGAG 480  
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaIleTrpValIleGluGluLys 159

QY 481 GCCTTCAGCGCGCGAGGTATCCCATGTTTCAACCGCGCTGAAGCGAGGCGCGCAGCGC 540  
Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179

QY 541 GACCTGAACACGATGTTGAACACCGTGGCGCGCGCAGCGCGCGCATCGAGTGTGAAG 600  
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199

QY 601 GACACCATCAACGAGAGAGCGCGCGCGAGTGGGACCGCGCTGCACCGCGCGCGCGC 660  
Db 200 AspThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisAlaGlyPro 219

QY 661 ATCGCGCGCGCGCGCAGATGCGCGAGCGCGCGCGCGCAGCAGCATCGCGCGCAGCAGC 720  
Db 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239

QY 721 CTGCAAGAGCAGATCCCTGATGATGACCAACCGCGCGCGCGCGCGCGCGCATCTAC 780  
Db 240 LeuGlnGluGlnIleAlaIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTrp 259

QY 781 AAGCGGTGATCATCTGGCGCTGAACAAGATCGTGGCGGATGTACAGCGCGTGTGAGCATC 840  
Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTrpSerProValSerIle 279

QY 841 CTGACATCAAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
Db 280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTrpValAspArgPhePheLys 299

QY 901 ACCGTGCGCGCGCGCAGAGCAGCAGCAGAGGTGAAGAACTGTGATGACCGACCGCTGTG 960  
Db 300 ThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 319

QY 961 GTGGAAGAGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1020  
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339

QY 1021 CTGGAAGAGATGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080

Db 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyProGlyHisLysAlaArgVal 359

QY 1081 CTGGCGGAGCGGATGAGCCAGCGCAACAC--AGCGTATGATGACAGAGCAACTTC 1137  
Db 360 LeuAlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPhe 379

QY 1138 AAGGCGCGCGCGCGCATCGTCAAGTGTCTCACTGCGCGCAAGAGGCGCGCATGCGCGC 1197  
Db 380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

QY 1198 AACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1257  
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

QY 1258 AAGACTGCACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1317  
Db 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439

QY 1318 CGCGCGCGCGCACTTCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
Db 440 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProAlaGlnSerPhe 459

QY 1378 CGCTTCGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437  
Db 460 ArgPheGluGluThrThrProAlaProLysGlnGluProIleGluArgGluProLeuThr 479

QY 1438 AGCCTGAAGAGCTGTTCGCGCAACGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1476  
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

## RESULT 4

US-10-332-413-8  
; Sequence 8, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swanstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdool  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 631  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-10-332-413-8

## Alignment Scores:

Pred. No.:	2.69e-120	Length:	631
Score:	2435.00	Matches:	458
Percent Similarity:	96.96%	Conservative:	20
Best Local Similarity:	92.90%	Mismatches:	13
Query Match:	87.84%	Indels:	2
DB:	16	Gaps:	2

US-09-475-704A-3 (1-1479) x US-10-332-413-8 (1-631)

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60

```
Db      31 MetAlaIaArgAlaSerIleLeuArgGluLysLeuAspLysTrpGluLysIleArg 50
QY      61 CTGCGCCCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGCCAGCCGCGAG 120
Db      51 LeuArgProGlyLysLysVshIstYkMetLeuLysHsiIleValTrpAlaSerArgGlu 70
QY      121 CTGAGAAGTTCCGCTGAACCCCGGCTGTGAGACGAGCGGCTGCAAGCAGATC 180
Db      71 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluLysSerLeuTyraAsn 90
QY      181 ATCCGCGAGCTGCACCCCGGCTGACAGACCGGACGAGAGTGAAGCCTGTTCAC 240
Db      91 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLysSerLeuTyraAsn 110
QY      241 ACCGTGGCCACCTGTACTGCGTGACAGAGATCGAGTCCGCGACCAAGAGGCC 300
Db      111 ThrValAlaThrLeuTyrcysValHisGluLysIleGluValArgAspThrLysGluAla 130
QY      301 CTGACAAGATCGAGGAGGACAGAACTGCGACGAGATCCAGACGCGCGAGGCC 360
Db      131 LeuAspLysIleGluGluGlnAsnLysCysGlnGlnLysThrGlnAlaLysAla 150
QY      361 GCCGACAAAGGCAAGGTAGCCAGAACTACCCCATCGTGACAACTGACAGGGCCAGATG 420
Db      151 AlaAsp---GlyLysValSerGlnAsnTyrrProIleValGlnAsnLeuGlnGlyMet 169
QY      421 GTGCACCAAGGCCATCAGCCCCCGCACCTGTGAAGCCTGGGTGAAGTATCGAGAGAG 480
Db      170 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 189
QY      481 GCCTTCAGCCCCGAGGTATCCCATGTTCACCGCCCTGAGCGAGGGGCCACCCCGCAG 540
Db      190 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 209
QY      541 GACCTGAACACGATGTTGAACACCGTGGGCGGCCACAGGCCCGCCATGACATGCTGAAG 600
Db      210 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 229
QY      601 GACACCATCAACGAGGAGGCCCGCGAGTGGGACCGCGGTGCACCCCGTGACGCGCGGCC 660
Db      230 AspThrIleAsnGluGluAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 249
QY      661 ATGCGCCCGCGGACAGATGGCGGAGCGCCGCGGACGACATCGCCGGGACCAACGACAC 720
Db      250 IleAlaProGlyGlnMetArgGluProArgLysSerAspIleAlaGlyThrThrSerThr 269
QY      721 CTGCAGAGCAGATCGCTGTGATGACCAACACCCCCCATCCCGTGGGCGACATCTAC 780
Db      270 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyrr 289
QY      781 AAGCGGTGATCATCTGTGGGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 840
Db      290 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrrSerProValSerIle 309
QY      841 CTGACATCAAGCAGGCGCCCAAGGAGCCTTCGCGCACTACGTGACCGCTTCTTCAAG 900
Db      310 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrrValAspArgPhePheLys 329
QY      901 ACCCTGGCGCGCGAGCAGACACCCAGAGGTGAAGAACTGGATGACCGACACCCCTGCTG 960
Db      330 ThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 349
QY      961 GTGCAGAACGCCAACCCCGACTGCAAGACATCTGCGCGCTCTCGGCCCGCGCGCGCAG 1020
Db      350 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 369
QY      1021 CTGAGAGATGATGACCGCTGCGACGCGGTGGCGCGGCCAGCCACAGAGCCCGCGCTG 1080
Db      370 LeuGluGluMetMetThrAlaCysGlnGlyValGlyLysProGlyHisLysAlaArgVal 389
QY      1081 CTGCGCGAGGCGATGAGCGAGGCCAACAC--AGCGTATGATGACAGAGCAACTTC 1137
```

```
Db      390 LeuAlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPhe 409
QY      1138 AAGGCGCCCGCGGCGCATCGTCAAGTGTCTCACTGCGGCAAGAGGGCCACATCGCCCGC 1197
Db      410 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 429
QY      1198 AACTGCGCGCGCCCGCGCAAGAGGGGTGCTGGAAGTGGCGCAAGAGGGCCACCAAGATG 1257
Db      430 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 449
QY      1258 AAGGATGCAACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCTGGGCCAGCCACAAGGGC 1317
Db      450 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 469
QY      1318 CGCCCGGCAACTTCTGACAGAGCGCGCCGAGCCACCGCCCGCGCGAGAGCTTC 1377
Db      470 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProAlaGluSerPhe 489
QY      1378 CGCTTCAGAGAGACCAACCCCGGCGCAAGCAGAGAGCAAGACCGCGAGACCTGACC 1437
Db      490 ArgPheGluGluThrThrProAlaProLysGlnGluProIleGluArgGluProLeuThr 509
QY      1438 AGCTGAAGAGCCTGTTCGCAACGACCCCTGAGCCAG 1476
Db      510 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 522
```

## RESULT 5

```
US-10-339-217-110
: Sequence 110, Application US/10339217
: Publication No. US20030198648A1
: GENERAL INFORMATION:
: APPLICANT: Buechter, Douglas
: APPLICANT: Hou, Xiaohong
: APPLICANT: Marlor, Christopher W.
: APPLICANT: Rice, William G.
: APPLICANT: Yang, Wengang
: TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
: TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA
: FILE REFERENCE: 111021.143(ACH-US1)
: CURRENT APPLICATION NUMBER: US/10/339,217
: PRIOR FILING DATE: 2003-01-09
: PRIOR APPLICATION NUMBER: US 60/347,369
: NUMBER OF SEQ ID NOS: 144
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 110
: LENGTH: 492
: TYPE: PRT
: ORGANISM: HIV-1
US-10-339-217-110
```

## Alignment Scores:

Pred. No.:	1.14e-119	Length:	492
Score:	2423.00	Matches:	451
Percent Similarity:	96.96%	Conservative:	27
Best Local Similarity:	91.48%	Mismatches:	13
Query Match:	87.41%	Indels:	2
DB:	14	Gaps:	2

US-09-475-704A-3 (1-1479) x US-10-339-217-110 (1-492)

```
QY      1 ATGGGCGCCGCGCGCAGCATCTGTGCGCGCGGCAAGCTGACGCGCTGGAGCGCATCGCG 60
Db      1 MetGlyAlaArgAlaSerIleLeuArgGlyGluLysLeuAspThrTrpGluLysIleArg 20
QY      61 CTGCGCGCCGCGGCAAGAGTGTACTATGATGAAGCACTGTGTGGGCGCAGCGCGAG 120
Db      21 LeuArgProGlyGlyLysLysCysTyrrMetLeuLysHisIleValTrpAlaSerArgGlu 40
QY      121 CTGAGAAGTTGCGCCTGAACCCCGCGCTGTGAGACCAAGAGGGCTGCAAGCAGATC 180
Db      41 LeuGluArgPheSerLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
```

```
QY 181 ATCCGCCAGCTGCACCCCGCCCTGCAGACCGGCGAGGAGCTGAAGACCTGTTCAAC 240
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGluLeuLysSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGGTGCACGAGAAAGATCGAGTCCGGCACCAAGAGAGCC 300
Db 81 ThrValAlaThrLeuPheCysValHisGlyLysIleAlaValArgAspThrLysGluAla 100
QY 301 CTGCACAGATCGAGAGAGAGAGAGAGAACAACTGCCAGAGAGATCCAGAGAGCCGAGGCC 360
Db 101 LeuAspLysIleGluGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120
QY 361 GCCGACAGGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAAGAACCTGCAGAGGCCAGATG 420
Db 121 AlaAsp---GlyThrValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 139
QY 421 GTGCACCAAGCCCATCAGCCCGCCGACCCCTGGAACGCGCTGGGTGAAGGTGATCGAGAGAG 480
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValIleGluGluLys 159
QY 481 GCCTTCAGCCCGCAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCCACCCCGCAG 540
Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179
QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCACCGCCCATGTCAGATGCTGAAG 600
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199
QY 601 GACACCATCAACGAGAGAGAGCGCCGCGAGTGGAGACCGCGCTGCACCCCGCTGCACGCGGCCCC 660
Db 200 AspThrIleAsnGlnGluAlaAlaGluTyrPaspArgLeuHisProValHisAlaGlyPro 219
QY 661 ATCGCCCCCGCCAGATGCGCGGAGGCGCGCGCGGCGAGCGACATCGCGCGCACCGACGACC 720
Db 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239
QY 721 CTGCAGAGAGAGATCGCGCTGGATGACCAACCCCGCATCCCGGTGGGCGACATCTAC 780
Db 240 LeuGlnGlnGlnIleAlaTyrMetThrAsnAsnProProValProValGlyAspIleTyr 259
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 840
Db 260 LysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279
QY 841 CTGCACATCAAGCAGGAGCGCCCAAGAGAGCCCTTCCGGAAGTACGTGACCGCTTCTTCAAG 900
Db 280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299
QY 901 ACCCTGCGCGCGCAGAGAGAGACCCAGAGGTGAAGAACTGGATGACCGACACCCCTGCTG 960
Db 300 ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTyrMetThrAspThrLeuLeu 319
QY 961 CTGCAGAACCGCAACCCCGACTGCAGAACCATCTCGCGCGCTCTGGGCGCCCGCGCGCAGC 1020
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339
QY 1021 CTGCAGAGAGATGATGACCGCTGCGAGGGCGTGGGCGGCGCCAGCGCAAGGCGCGGTG 1080
Db 340 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359
QY 1081 CTGGCCGAGGCGGATGAGCCAGGCC---AACACCAAGCGTGATGATGCAGAGAGCACTTC 1137
Db 360 LeuAlaGluAlaMetSerGlnThrAsnAsnAlaAsnIleMetMetGlnArgSerAsnPhe 379
QY 1138 AAGGGCCCCCGCGCGCATCTGCAAGTGTCTCAACTGCGGCAAGAGGGGCCACATCGCCGCG 1197
Db 380 LysGlyProArgArgIleIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 399
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGGCAAGAGGGGCCACAGATG 1257
Db 400 AsnCysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGluGlyHisGlnMet 419
QY 1258 AAGGACTGCACCGAGCGCGCAAGGCCAACTTCTGGGCAAGATCTGGCCAGGCCCAAGGGC 1317
```

```
Db 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTyrProSerHisLysGly 439
QY 1318 CGCCCCGGCAACTTCTTCAGAGCCGCGCCGAGCCACCGCCCCCGCCGAGAGCTTC 1377
Db 440 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 459
QY 1378 CGCTTCGAGAGAGACACCCCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
Db 460 ArgPheGluGluThrThrProAlaProLysGlnGluProArgGluArgGluProLeuThr 479
QY 1438 AGCCTGAAGAGCCTGTTCCGCAACGACCCCTGAGCCAG 1476
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 6
US-10-339-217-109
; Sequence 109, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; FILE REFERENCE: 111021.143 (ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 109
; LENGTH: 492
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-109

Alignment Scores:
Pred. No.: 2.66e-119 Length: 492
Score: 2416.00 Matches: 455
Percent Similarity: 95.74% Conservative: 17
Best Local Similarity: 92.29% Mismatches: 19
Query Match: 87.16% Indels: 2
DB: 14 Gaps: 2

US-09-475-704A-3 (1-1479) x US-10-339-217-109 (1-492)
QY 1 ATGGGCGCGCGCCAGCATCCTGCGCGGCGCAAGCTGAGCGCCTGGAGCGCATCCGC 60
Db 1 MetGlyAlaSerAlaSerIleLeuArgGlyGlyLysLeuAspLysTyrGluLysIleArg 20
QY 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGCGCAGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysLysTyrArgLeuLysHisLeuValTyrAlaSerArgGlu 40
QY 121 CTGAGAAAGTTCGCCCTGAACCCCGGCTGCTGAGAGACGAGCGAGGCTGAAGAGCTGTTCAAC 180
Db 41 LeuGluArgPheAlaLeuAsnSerGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60
QY 181 ATCCGCAAGCTGCACCCCGCCCTGCAGACCGGCGAGCGAGAGCTGAAGAGCTGTTCAAC 240
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGluLeuLysSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGGTGCACGAGAAAGATCGAGTCCGGCACCAAGAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluValArgAspThrLysGluAla 100
QY 301 CTGCACAGATCGAGAGAGAGAGAGAACAACTGCCAGAGAGATCCAGAGAGCCGAGGCC 360
Db 101 LeuAspLysIleGluGlnGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysGlu 120
```

```
QY 361 GCCGACAAGGCAAGGTGAGCCAGAACTAATCCCATCTGTCAGAACTGAGGCGGCAAGATG 420
Db 121 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMet 139
QY 421 GTGCACCAAGCCATCAGCCCCCGACCCCTGAAGCGCTGGTGAAGGTGATCGAGAGAG 480
Db 140 ValHisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCGAG 540
Db 160 AlaPheSerProGluValIlePrometPheThrAlaLeuSerGluGlyAlaThrProGln 179
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGGCGCCACGAGCGCCCATGAGATGCTGAAG 600
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199
QY 601 GACACCATCAACGAGGAGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCC 660
Db 200 AspThrIleAsnGluGlnAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 219
QY 661 ATCGCCCCCGGCGCAGATGCGCGGAGCGCGCGGCGGACGACATCGCGCGGACGACGACC 720
Db 220 ValAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239
QY 721 CTGCAGAGACGATCGCCTGGATGACCAACGACACCCCATCCCGTGGCGGACATCTAC 780
Db 240 LeuGlnGluGlnIleThrTrpMetThrSerAsnProProValProValGlyAspIleTyr 259
QY 781 AAGCGGTGATCATCTCTGGGCTGTAACAAGATCGTGGCATGTACAGCCCGTGAAGATC 840
Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279
QY 841 CTGCACATCAAGCAGGCGCCCAAGAGACCTTCGCGCACTACGTGACCGCTTCTTCAAG 900
Db 280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299
QY 901 ACCCTGCGCGCGCGAGACGACGACCCGAGAGTGAAGAACTGATGACGACACCTGCTG 960
Db 300 ValLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319
QY 961 GTGCAGAACGCAACCCCGACTGCACAGACCATCTCGCGCGCTCTCGGCGCGCGCGGCAAC 1020
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProAlaAlaSer 339
QY 1021 CTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGGCGCCGACGACAGGCGCGGTG 1080
Db 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359
QY 1081 CTGCGCGAGGCGGATGAGCCAGGCGCAAC--ACCAGCTGATGATGACGAAGAGCAACTTC 1137
Db 360 LeuAlaGluAlaMetSerGlnAlaAsnThrThrAsnIleMetMetGlnLysSerAsnPhe 379
QY 1138 AAGGCGCGCGCGCGCATCTGTCAAGTGTCTCAACTGCGGCAAGAGGCGCACATGCGCGC 1197
Db 380 LysGlyProArgArgThrValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLys 399
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGCGTCTGGAAGTGGCGGCAAGAGGCGCCACGATG 1257
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysGlyLysGluGlyHisGlnMet 419
QY 1258 AAGGACTGCACCGAGCGCGCAGGCGCAACTTCTGGGCAAGATCTGGCGGACGACAGAGGC 1317
Db 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439
QY 1318 CGCGCGCGCAACTTCTGTCAGAGCGCGCGCGGCGGCGGCGCGCGCGGAGAGCTTC 1377
Db 440 ArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPhe 459
QY 1378 CGCTTCGAGGAGACCAACCGCGCGCGGCGGCAAGAGGAGCAAGGACCGGAGACCTGACC 1437
Db 460 ArgPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuThr 479
```

```
QY 1438 AGCTGAAGAGCCTGTTCCGCAACGACCCCTGAGCCAG 1476
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 7
US-10-339-217-140
; Sequence 140, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140
; LENGTH: 491
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-140
```

```
Alignment Scores:
Pred. No.: 2.82e-119 Length: 491
Score: 2415.50 Matches: 450
Percent Similarity: 96.75% Conservative: 26
Best Local Similarity: 91.46% Mismatches: 15
Query Match: 87.14% Indels: 1
DB: 14 Gaps: 1
```

US-09-475-704A-3 (1-1479) x US-10-339-217-140 (1-491)

```
QY 1 ATGGGCGCGCGCGCGGACATCTCTGGCGCGGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20
QY 61 CTGGCGCGCGCGGCAAGAGTGTCTACATGATGAAGCACTGTGTGGGCGGCGGCGGAG 120
Db 21 LeuArgProGlyGlyLysLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40
QY 121 CTGAGAGAGTTCGCGCTGAACCCCGCGCTGTGAGACGAGCGAGGCTGCAAGCAGATC 180
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60
QY 181 ATCGCGCAGCTGACCCCGCGCTGAGACCGGCGGAGGAGGAGCTGAAGAGCTGTTC AAC 240
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCGACCTGTACTGCGTGCACGAGAAGATCGAGGTCCGCGACACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGluIleGluValArgAspThrLysGluAla 100
QY 301 CTGACAAGATCGAGGAGGAGGAGCAACAAGTGCAGACGAGAGATCCAGAGCGCGAGGCC 360
Db 101 LeuAspArgIleGluGluGluGlnAsnLysSerGlnLysThrGlnGlnAlaAsnGlu 120
QY 361 GCCGACAAGGCGCAAGGTGAGCGCAACTACCCCATCTGTCAGAACTGCGAGGCGGAGATG 420
Db 121 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMet 139
QY 421 GTGCACCAAGCCATCAGCCCCCGGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAG 480
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCGGCGGCGGAG 540
```

```
Db      160 AlapheserProgluValIlePrometPheThrAlaleuSerGluGlyAlaThrProGln 179
QY      541 GACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCCCGCATGCAGATGCTGAAG 600
Db      180 AspleuanThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLys 199
QY      601 GACACCATCAACGAGGAGGCGCCGAGTGGGAGCGCGGTGCACCCCGTGCACCGCGGCCCC 660
Db      200 AspThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisProValHisAlaGlyPro 219
QY      661 ATGCCCCCGCCGAGATGGCGAGAGCCCGCGGAGCGAGCATGCGCGGACCAACAGCAGACC 720
Db      220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerSer 239
QY      721 CTGCAGGACGATCGCTGGATGACCAACCCCGCATCCCGTGGGCGGACATCTAC 780
Db      240 LeuGlnGluGlnIleAlaThrMetThrGlyAsnProProValProValGlyAspIleTyr 259
QY      781 AAGCGTGGATCATCTCTGGGCTGAACAGATCGTGGGATGTACAGCCCGTGAAGCATC 840
Db      260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279
QY      841 CTGCACATCAAGCAGGCGCCCAAGGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900
Db      280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299
QY      901 ACCCTGGCGCGCCGAGCAGACCCCAAGAGGTGAAGAACTGATGACGACACCCCTGCTG 960
Db      300 ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319
QY      961 GTGCAGAACGCCAACCCCGACTGCAGAACCATCTCTGCGCGCTCTCGCGCCCGCGCCAGC 1020
Db      320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 339
QY      1021 CTGCAGAGATGATGACCGCTGCGCAGGCGGTGGCGGCCCAAGGCCCAAGGCCCGCTG 1080
Db      340 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359
QY      1081 CTGGCCGAGGCGGATGAGCCAGGCCCAACACCAACCGTGTATGATGACAGAGCAACTTCAAG 1140
Db      360 LeuAlaGluAlaMetSerGlnThrAsnSerThrIleLeuMetGlnArgSerAsnPheLys 379
QY      1141 GGGCCCGCGCGCATCTCAAGTGTCTCAACTGCGGCAAGAGGAGGCCCATCGCCGCAAC 1200
Db      380 GlyProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLysAsn 399
QY      1201 TGCCGCGCCCGCCGCAAGAGGCTGTGAGTGTGAGTGTGAGGAGGAGGCCACCAAGTGAAG 1260
Db      400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 419
QY      1261 GACTGCACCGAGCGCCAGGCGCAACTTCTTGGGCAAGATCTGCGCCAGCCACCAAGGCGCGC 1320
Db      420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439
QY      1321 CCGGCACTTCTCTGAGAGCGCGCCCGAGCGCCACCGCCCGCCCGAGAGCTTCCGC 1380
Db      440 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 459
QY      1381 TTGCAGGAGACACCCCGCGGAGAGCAGAGCAAGAGCAAGCAGAGCAGACCTGACAGC 1440
Db      460 PheGluGluThrThrProAlaLeuGlnGlnGlyProLysAspArgGluProLeuThrSer 479
QY      1441 CTGAAGAGCCTGTTGGCAACGACCCCTGAGCCAG 1476
Db      480 LeuArgSerLeuPheGlySerAspProLeuSerGln 491
```

RESULT 8  
US-10-339-217-104

; Sequence 104, Application US/10339217  
; Publication No. US20030198648A1  
; GENERAL INFORMATION:  
; APPLICANT: Buechter, Douglas  
; APPLICANT: Hou, Xiaohong

```
; APPLICANT: Marlor, Christopher W.  
; APPLICANT: Rice, William G.  
; APPLICANT: Yang, Wengang  
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of  
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA  
; FILE REFERENCE: 111021.143 (ACH-US1)  
; CURRENT APPLICATION NUMBER: US/10/339,217  
; PRIOR APPLICATION NUMBER: 2003-01-09  
; PRIOR FILING DATE: 2002-01-11  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 104  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: HIV-1  
US-10-339-217-104  
  
Alignment Scores:  
Pred. No.: 7,45e-119 Length: 491  
Score: 2407.50 Matches: 449  
Percent Similarity: 96.34% Conservative: 25  
Best Local Similarity: 91.26% Mismatches: 17  
Query Match: 86.85% Indels: 1  
DB: 14 Gaps: 1  
  
US-09-475-704A-3 (1-1479) x US-10-339-217-104 (1-491)  
  
QY      1 ATGGCGCGCCGCGCCAGCATCTCTGCGCGCGCAAGCTGACGCTGGAGCGCATCCGC 60  
Db      1 MetGlyAlaArgAlaSerIleLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20  
QY      61 CTGCGCGCCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGCGCAGCCGCGAG 120  
Db      21 LeuArgProGlyGlyArgLysHisTyrMetLeuLysHisIleuValTrpAlaSerArgGlu 40  
QY      121 CTGCAGAGTTCGCCCTGAACCCCGCTGTGAGACGACGAGGCTGGAAGCAGATC 180  
Db      41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysLysGlnIle 60  
QY      181 ATCCGCGAGCTGCACCCCGCTGCAGACCGCGCAGAGAGCTGAAGCCTGTTCAAC 240  
Db      61 IleLysGlnLeuHisProAlaLeuLysThrGlyThrGluLysLeuArgSerLeuTyrAsn 80  
QY      241 ACCGTGCGCACTTCTACTGCGTGCAGAGAGATCGAGGTCCGCGACCAAGAGGCC 300  
Db      81 ThrValAlaThrLeuTyrCysValHisGlnAsnIleGluValArgAspThrLysGluAla 100  
QY      301 CTGCACCAAGATCGAGAGAGCAGAACAGTGCAGAGATCCAGAGCGCGAGGCC 360  
Db      101 LeuAspLysIleGluGluGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120  
QY      361 GCCGACAAGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACCTGCAAGGCCAGATG 420  
Db      121 AlaAspGluGly--ValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 139  
QY      421 GTGCACCAAGCCATCAGCCCGCGCACCTGAAGCCTGGGTGAAGTGAAGTGAAGAGAG 480  
Db      140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159  
QY      481 GCCTTCAGCCCGAGGTGATCCCATGTTCACCGCCTGAGCGAGGCGCCACCCCGCAG 540  
Db      160 AlapheserProgluValIlePrometPheThrAlaleuSerGluGlyAlaThrProGln 179  
QY      541 GACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCCCGCATGCAGATGCTGAAG 600  
Db      180 AspleuanThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLys 199  
QY      601 GACACCATCAACGAGGAGCGCCGAGTGGGAGCGCGGTGCACCCCGTGCACCGCGGCCCC 660  
Db      200 AspThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisProValHisAlaGlyPro 219  
QY      661 ATGCCCCCGCCGAGATGGCGAGAGCCCGCGGAGCGAGCATGCGCGGACCAACAGCAGACC 720
```

```
Db      220 AlaAlaProGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr
QY      721 CTGCAGAGCAGATCGCTTGATGACCAACCCCGCATCCCGTGGCGCAGATCTAC
      240 LeuGlnGlnGlnIleAlaTrpMetThrGlyAsnProProValProValGlyAspIleTyr
QY      781 AAGCGGTGATCATCTGGGCTTGAACAAGATCGTGGAGTATACAGCCCGTGAGCATC
      260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle
QY      841 CTGCACATCAAGAGGCCCCAAGAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAG
      280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys
QY      901 ACCCTGCGCGCCGAGCAGACCCCAAGAGGTGAAGAATGATGACCGACACCCCTGCTG
      300 ValLeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu
QY      961 GTGCAGAACGCCAACCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCGCGCGCAGC
      320 IleGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAlaAlaSer
QY      1021 CTGAGAGAGATGATGACCGCTGCAAGGCGGTGGCGGCCCAAGCCCAAGGCCGCGTG
      340 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgVal
QY      1081 CTGCGCGAGGCGATGACCCAGCCCAACCAAGCGTGATGATGACAGAAGCACTTCAAG
      360 LeuAlaGlnAlaMetSerGlnAlaAsnSerAsnIleMetMetGlnArgSerAsnPheLys
QY      1141 GGGCCCCGGCGCATCTCTCAAGTGTCTCAACTGCGGCAAGAGGGCCCATCGCCCGCAAC
      380 GlySerLysArgIleValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaArgAsn
QY      1201 TGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGTGCGCAAGAGGGCCACAGATGAAG
      400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGlnGlyHisGlnMetLys
QY      1261 GACTGCACCGAGCGCCAGGCCCACTTCTTGGCAAGATCTGGCCAGCCACCAAGGGCGCC
      420 AspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg
QY      1321 CCGCGCAACTTCTTGCAGAGCGCGCCCGAGCCACCGCCCGCCGAGAGCTTCCGC
      440 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg
QY      1381 TTCGAGGAGAGCACCACCGCGCCGAGAGCAGAGAGCAGAGCAGCCGCGAGACCTGACCAAC
      460 PheGlnGlnThrThrProAlaProLysGlnGlnSerLysAspArgGluProLeuIleSer
QY      1441 CTGAAGAGCCTGTTGGCAACGAGCCCTGAGCCAG 1476
      480 LeuLysSerLeuPheGlnSerAspProSerSerGln 491
RESULT 9
US-10-339-217-143
; Sequence 143, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 143
; LENGTH: 491
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-143

Alignment Scores:
Pred. No.:      2,83e-118      Length:      491
Score:          2396.50      Matches:      449
Percent Similarity: 95.53%      Conservative: 21
Best Local Similarity: 91.26%      Mismatches: 21
Query Match:    86.45%      Indels:      1
DB:             14      Gaps:      1

US-09-475-704A-3 (1-1479) x US-10-339-217-143 (1-491)

QY      1 ATGGCGCGCGCGCGCAGCATCTCGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60
      1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20
QY      61 CTGCGCGCGCGCGCGCAAGAAGTGTACATGATGAGCAACCTGTGTGGCGCCAGCGCGAG 120
      21 LeuArgProGlyLysLysHisLysTrpMetIleLysHisLeuValTrpAlaSerArgGlu 40
QY      121 CTGGAAGAAGTTCGCTTGAACCCCGCTGCTGAGACACCGAGCGGCTGCAAGCAGATC 180
      41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysLysGlnIle 60
QY      181 ATCGGCCAGCTGCAACCCCGCCCTGACAGACCGGCAAGAGAGCTGAAGAAGCTGTTCAC 240
      61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGluLeuArgSerLeuHisAsn 80
QY      241 ACCGTGGCCACCCCTGTACTGCGTGAACGAGAGATCGAGTCCGCGCACACCAAGAGGCC 300
      81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluIleArgAspThrLysGlnAla 100
QY      301 CTGCACAAGATCGAGAGAGAGCAACAAGTCCGCAAGAGATCCAGACGAGCGCGAGCC 360
      101 LeuAspLysIleGlnGlnGlnGlnLysSerGlnGlnLysThrGlnGlnAlaLysGlu 120
QY      361 GCCCAACAAGGCGCAAGGTGAGCCCAACTACCCCATCTGTCAAGAACTGCAAGGCGCCAGATG 420
      121 AlaAsp--GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 139
QY      421 GTGCACCAAGGCCATCAGCCCGCCGACCTGAACCGCTGGGTGAAGTGATCGAGAGAAAG 480
      140 ValHisGlnAlaLeuSerProArgThrLeuAsnAlaTrpValLysValIleGlnGluLys 159
QY      481 GCCTTCAGCCCGCGAGGTGATCCCATGTTCAACCGCGCTGAGCGAGGCGCCACCCCCAG 540
      160 AlaPheSerProGluIleIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 179
QY      541 GACCTGAACACGATGTTGAACACCGTGGCGGCGCCACGAGCGCGCATGACAGATGCTGAAG 600
      180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199
QY      601 GACACCATCAACAGAGAGAGCGCGCGAGTGGAGCGCGGTGCACCCCGTGCACGCGCGGCC 660
      200 AspThrIleAsnGlnGlnAlaAlaGlnTrpAspArgLeuHisProAlaGlnAlaGlyPro 219
QY      661 ATGCGCGCGCGCGCAGATGCGCGAGCGCCCGCGGAGCGAGACATCGCGGCGCACCAAGCACC 720
      220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239
QY      721 CTGCAGAGCAGATCGCTTGATGACCAACCCCGCATCCCGTGGCGCAGATCTAC 780
      240 LeuGlnGlnGlnIleAlaTrpMetThrGlyAsnProProValProValGlyLysIleTyr 259
QY      781 AAGCGGTGATCATCTGGGCTTGAACAAGATCGTGGAGTATACAGCCCGTGAGCATC 840
      260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279
```



QY	1006	GGCCCCGGCCGCA	CGCTGAGAGATGATGACCGCCTG	CCAGGGCGTGGCGGCCCA	GC	1065	
			:::				
Db	340	GlyProGlyAlaThr	LeuGluMetMetThr	AlaCysGlnGlyValGly	ProSer	359	
QY	1066	CACAAGGCCCGCTG	CTGGCCGAGCGATGAGCCAGCA	ACCACCGCTGATGATGAC		1125	
Db	360	HisIysAlaArgVal	LeuAlaGluAlaMetSer	GlnThrAsnSerAsn	LeuMetMetGln	379	
QY	1126	AAGAGCAACTCAAG	GGCCCCGGCGCATGTC	CAAGTGTTC	CACTGCGGCAAGAGGCGC	1185	
Db	380	AsnSerAsnPhel	ysGlySerArgArgIle	ValIysCysPheAsn	CysGlyIysValGly	399	
QY	1186	CACATCGCCCGCA	ACTGCCGCGCCCCCGCA	AGAGGGCTGCTG	GAAGTGGCGCAAGAG	1245	
Db	400	HisIleAlaArgAsn	CysArgAlaProArg	IleIysGlyCys	TrpIysCysGlyIysGln	419	
QY	1246	GGCCACCAATGAAG	ACTGCACCGAGCGCCAGGCCA	ACTTCTG	GGCAGAATCTGGCCC	1305	
Db	420	GlyHisGlnMet	LysAspCysThrGlu	ArgGlnAlaAsn	PhelLeuGlyArgIle	TrpPro	439
QY	1306	AGCCACAGGGCGCC	CGCCGCACTTCTCTG	CAAGCGCCCCCGAG	CCCA	CGCCCCCCC	1365
Db	440	SerHisIysGly	ArgProGlyAsn	PhelLeuGlnSer	ArgProGluPro	ThrAlaProPro	459
QY	1366	GCCGAGAGCTTCCG	CTTCGAGAGACCA	CCCCCGGCCA	GAGCAGAGCAAGCAAG	AGCAGC	1425
Db	460	AlaGluSerPhe	ArgPheGlnGluIle	ThrProValPro	LysGlnGluPro	LysAspArg	479
QY	1426	GAGACCTGACCA	GCCTGAAGAGCTGT	TCGGCAAGAC	CCCCCTGAGCCAG	1476	
Db	480	GluProLeuThr	SerLeuIysSer	LeuPheGlySer	AspProLeuSer	Gln	496

```

RESULT 11
US-10-339-217-106
; Sequence 106, Application US/10339217
; Publication No. US20030198648A1
;
GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
;
NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
;
LENGTH: 487
;
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-106

```

Alignment Scores:	
Pred. No.:	1.74e-117
Score:	2381.50
Percent Similarity:	95.94%
Best Local Similarity:	91.08%
Query Match:	85.91%
DB:	14
Length:	487
Matches:	449
Conservative:	24
Mismatches:	13
Indels:	7
Gaps:	2

US-09-475-704A-3 (1-1479) x US-10-339-217-106 (1-487)

QY 1 ATGGCGCCGCGCAGATCTCGCGCGCGGCAAGCTGGAACGCCTGGAGCGCATCCG 60  
||| ||| ||| :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 1 Metcylalaa rgala se rVal leu l ysc i ly l s yl s le u as p Thr t r p c i ua r g i le a r g 20  
  
QY 61 CTGCGCCCCGCGCGCAAGAAGTGTACATGATGAGCAGCCTGTGTGGGCCAGCGCGAG 120

D	b	21	LeuArgProGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGlu	40
Q	y	121	CTGGAGAAGTTGCGCCTGAACCCCGGCTGTGTGAGACCAAGGAGGCTGCAAGCATC	180
D	b	41	LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle	60
Q	y	181	ATCGCCAGCTGCACCCCGCCTGCAGACCCGAGCGAGAGGTGAAGCCTTTCAC	240
D	b	61	MetGlnGlnLeuGlnSerAlaLeuGlnThrGlyThrGluLeuArgSerLeuTyrAsn	80
Q	y	241	ACCGTGGCCACCTGTACTGCGTGCACGAGAATCGAGTCCGCGACACCAAGAGGCC	300
D	b	81	ThrValAlaThrLeuTyrCysValHisLysGlnIleAspValArgAspThrLysGluAla	100
Q	y	301	CTGGACAAGATCGAGGAGGAGGACAACAAGTGCACGACGAGAATCCAGAGCGGAGCC	360
D	b	101	LeuAspLysIleGluGluGluGlnAsnLysSerGlnGlnLysThrGlnIleAlaGluAla	120
Q	y	361	GCCGACAAGGGCAAGGTGAGCCGAACTACCCCATCGTGACAACCTGCAGGCGCATG	420
D	b	121	AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet	140
Q	y	421	GTGCACCAAGGCATCAGCCCCGCAACCTGAACGCTGGTGAAGTGATCGAGAGAAG	480
D	b	141	ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys	160
Q	y	481	GCGTTCAAGCCCGAGGTGATCCCATGTTCAAGCCCTGAGCGAGGGCGCAACCCCGAG	540
D	b	161	AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln	180
Q	y	541	GACCTGAACAAGATTGTTGAACAACGCTGGCGGCCACGAGCGCCATGCAGATGCTGAAG	600
D	b	181	AspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaIleMetGlnMetLeuLys	200
Q	y	601	GACACCATCAAGAGAGGCGCGGAGTGGAGCGCGTGCACCCCGTGACGCGCGGCC	660
D	b	201	AspThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisAlaGlyPro	220
Q	y	661	ATCGCCCCCGCGCATGCGCGAGCCCCGCGGACGACATCGCCGCGCAACCAAGACCC	720
D	b	221	IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr	240
Q	y	721	CTGCAGAGCAAGATCGCCTGATGACCAAGCAACCCCATCCCGTGGCGCATCTAC	780
D	b	241	LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyr	260
Q	y	781	AAGCGGTGATCATCTGGGCTGAACAAGATCGTCGGATGTACAGCCCCGTGAGCATC	840
D	b	261	LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle	280
Q	y	841	CTGCACATCAAGAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTCAAG	900
D	b	281	LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys	300
Q	y	901	ACCGTGGCGCGGACGACGACCAAGAGAGGTGAAGAATGTGATGACCGACCCCTGCTG	960
D	b	301	ThrLeuArgAlaGluGlnSerSerGlnIleValLysAsnTrpMetThrAspThrLeuLeu	320
Q	y	961	CTGCAGAACGCAACCCCGACTGCAAGACCATCTCGCGCTTCGCGCCGCGCGCCAGC	1020
D	b	321	ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr	340
Q	y	1021	CTGGAGGAGATGACCGCGCTGCCAGGCGGTGGGCGGCCCAAGCCACAAGCGCGGTG	1080
D	b	341	LeuGluGluMetMetThrAlaCysGlnGlyValGlyProGlyHisLysLeuAlaArgVal	360
Q	y	1081	CTGGCCGAGGCGATGAGCCAGGCCAACACCAAGCGTGATGTACAGAAGACAATTCAAG	1140
D	b	361	LeuAlaGluAlaMetSerGlnAlaAsnThrAsnIleMetMetGlnLysSerAsnPheLys	380
Q	y	1141	GCGCCCGCGCGCATGTCAGTGTCTCACTGCGGCAAGAGGGCCACATGCGCCGAAC	1200



```
|||||
Db      440 ProGlyAsnPhLeuGlnAsnArgProGluProThrAlaProProAlaGluProThrAla 459
QY      1360 CCCCCCGAGAGCTTCGCTTCGAGAGACACCCCGCCAGAGAGAGAGAG 1419
Db      460 ProProAlaGluSerPheArgPheGluGluThrThrProValProLysArgGlu--Lys 478
QY      1420 GACCGGAGACCTTGACACGCTGAAGAGCTGTTCGCAACGACCCCTGAGCCAG 1476
Db      479 GluArgGluProLeuThrSerLeuLysSerLeuPheGlyAsnAspProSerSerGln 497

RESULT 13
US-10-339-217-105
; Sequence 105, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105
; LENGTH: 508
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-105

Alignment Scores:
Pred. No.:      1.64e-115      Length:      508
Score:          2344.00        Matches:     446
Percent Similarity: 93.32%      Conservative: 29
Best Local Similarity: 87.62%      Mismatches:  16
Query Match:     84.56%        Indels:      18
DB:              14            Gaps:         4

US-09-475-704A-3 (1-1479) x US-10-339-217-105 (1-508)
QY      1 ATGGGCGCCGCGCGGCGATCTGCGCGGCGGCAAGCTGAGCGCTGGGAGCGCATCCGC 60
Db      1 MetGlyAlaArgAlaSerIleLeuArgGlyThrLysLeuAspAlaTrpGluLysIleArg 20
QY      61 CTGCGCGCCGCGGCGGCAAGTGTATGATGAAGCACTGTGTGGCCAGCGCGAG 120
Db      21 LeuArgProGlyLysLysLysIleTyrMetLeuLysIleValTrpAlaSerArgGlu 40
QY      121 CTGAGAAGTTCGCTGAACCCCGGCTGTGAGACGAGCGAGGCTGCAAGAGATC 180
Db      41 LeuGluArgPheAlaLeuAsnProGlyLeuGluThrSerGluGlyCysLysGlnIle 60
QY      181 ATCCGCGAGCTGACACCCCGGCTGAGACCGGCGAGGAGAGCTGAAGCCTGTTCAG 240
Db      61 MetLysGlnLeuHisProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80
QY      241 ACCGTGGCCACCTGTACTGCGTGCAAGAGATGAGTCCGCGACACCAAGAGAGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHisGluAsnIleLysValArgAspThrLysGluAla 100
QY      301 CTGCAACAAGATCGAGAGAGAGACA-----AAGTGCACAGACAAGATCCAGAGGCC 354
Db      101 LeuAspLysIleGluGluGluGlnAsnLysIleLysSerGlnGlnLysThrGlnGlnAla 120
QY      355 GAGCGCGCGCGACAAAGGCGCAAGGTGAGCGACGACTACCCCATCTGTGACGAACCTG 414
Db      121 LysAlaAlaAspGlu---LysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 139
```

```
QY      415 CAGATGGTGACACCAAGCCATCAGCCCCCGACCCCTGAACGCTGGGTGAAGTGATCGAG 474
Db      140 GlnMetValHisGlnAsnLeuSerProArgThrLeuAsnAlaTrpValLysValIleGlu 159
QY      475 GAGAGGCTTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGGAGGCGCCACC 534
Db      160 GluLysAlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThr 179
QY      535 CCCCAGACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCGCCCATGACAGATG 594
Db      180 ProGlnAspLeuSerThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMet 199
QY      595 CTGAAGGACACCATCAACGAGAGAGCGCCGAGTGGGACCGGTCGACCCCGTGACAGCC 654
Db      200 LeuLysAspThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisAla 219
QY      655 GAGCCCATCGCCCCCGGCGAGATGGCGGAGCCCGGCGGAGAGACATCGCCGAGACACC 714
Db      220 GlyProMetAlaProGlyGlnLeuArgGluProArgGlySerAspIleAlaGlyThrThr 239
QY      715 AGCACCTTGACGAGAGATGCGCTGGATGACCAACACCCCCCATCCCGTGGGCGAC 774
Db      240 SerThrLeuArgGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAsp 259
QY      775 ATCTCAAGCGGTGATCATCTCGGCGCTGAACAAGATCGTGGCATGTACAGCCCGCTG 834
Db      260 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProVal 279
QY      835 AGCATCTTGACATCAACGAGAGGCGCCCAAGAGAGCCCTTCGCGCATACGTGACCGCTTC 894
Db      280 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 299
QY      895 TTCAGAACCCCTGCGCGCGGCGGAGACACCAAGAGAGTGAAGATGACCGACACCC 954
Db      300 PheLysAlaLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrGluThr 319
QY      955 CTGCTGTGTCAGAACCGCCCAACCCCGCATGCAAGACCATCTGCGCGCTCTCGGCCCGCC 1014
Db      320 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyIleGly 339
QY      1015 GCCAGCCTTGAGAGATGATGACCGCCTGCGAGGCGGTGGCGGCCCGCCACCAAGGCC 1074
Db      340 AlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAla 359
QY      1075 CGCGTGTGCGCGGCGGATGAGCCAGGCC--AACACCAAGCTGATGACAGAGAGCC 1131
Db      360 ArgValLeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetMetGlnArgSer 379
QY      1132 AACTTCAAGGGCGCGCGCATGCTCAAGTCTTCAACTGCGGCAAGAGGCGCACATC 1191
Db      380 AsnPheLysSerSerLysArgIleValLysCysSerAsnCysGlyLysGluGlyHisIle 399
QY      1192 GCCCGCAACTGCGCGGCGCCCGCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCAC 1251
Db      400 AlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHis 419
QY      1252 CAGATGAAGACTGCAACGAGCGGCGCAAGCTTCTGCGCAAGATCTGCGCCAGCGCAC 1311
Db      420 GlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHis 439
QY      1312 AAGGCGCGCGCGCAACTTCTGCGAG----- 1338
Db      440 LysGlyArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGlu 459
QY      1339 -----AGCGCGCGCGAGCCGACCGCCCGCGCGAGAGCTTCGCTTCGAGAG 1389
Db      460 SerPheArgAsnArgProGluProThrAlaProProAlaGluSerPheArgPheGluGlu 479
QY      1390 ACCACCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1449
Db      480 ThrThrProThrProLysGlnGluProLysAspArgAspProLeuThrSerLeuLysSer 499
```

QY 1450 CTGTTGGCAAGACCCCTGAGCCAG 1476  
Db 500 LeuphegLyserAspProserSerGln 508

RESULT 14

US-09-899-575-22  
; Sequence 22, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megede, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Estrelita Janse  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; CURRENT FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 09/475,704  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus  
US-09-899-575-22

Alignment Scores:

Pred. No.:	1.85e-115	Length:	502
Score:	2343.00	Matches:	445
Percent Similarity:	93.64%	Conservative:	26
Best Local Similarity:	88.47%	Mismatches:	20
Query Match:	84.52%	Indels:	12
DB:	10	Gaps:	4

US-09-475-704A-3 (1-1479) x US-09-899-575-22 (1-502)

QY 1 ATGGGGCGCCGCGCGCAGCATCTGCGCGCGCGCAAGCTGAGCGCCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGlyLysIleArg 20  
QY 61 CTGCGCCCGCGCGCGCAAGAGTGCTACATGATGAGCACTGTGTGGCCAGCCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGly 40  
QY 121 CTGGAAGAAGTTGCGCCTGAACCCCGCCTGTGAGACCGAGCGAGAGCTGAAGCCTGTTCAAC 180  
Db 41 LeuGlnGlyPheAlaLeuAsnProGlyLeuLeuGluThrAlaGlnGlyCysLysGlnIle 60  
QY 181 ATCCGCCAGCTGACCCCGCCTGAGACCGGACGAGGAGAGCTGAAGCCTGTTCAAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluValArgAspThrLysGlnAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGCAAGCAAGTGCGACGAGAAGATCGACGAGCGGAGCC 360  
Db 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120  
QY 361 GCGGACAAGGCGAGGTGAGCCAGAACTACCCCATCGTGACAGACCTGACGAGCGCAGATG 420  
Db 121 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 139  
QY 421 GTGCACCAAGGCCATGACCCCGCAGCCCTGAAACGCTGGTGAAGGTGATCGAGAGAAG 480  
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 159  
QY 481 GCCTTCAAGCCCGAGGTGATCCCCATGTTCAACCGCCTGAGCGAGGCGCCACCCCCAG 540  
Db 160 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 179

QY 541 GACCTGAACAGATGTTGAACACCGTGCGCGCCACAGCGCCCATGACAGATGCTGAAG 600  
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAACGAGAGCGCGCGAGTGGGACCGCGTGCAACCCCGTGACCGCGGCC 660  
Db 200 AspThrIleAsnGlnGlnAlaAlaGlnGluTrpAspArgLysHisProValGlnAlaGlyPro 219  
QY 661 ATCGCCCGCGCAGATGCGCGGAGCCCGCGCGCAGCGACATCGCCGCGCACCAAGCACC 720  
Db 220 ValAlaProGlyGlnMetArgAspProArgGlySerAspIleAlaGlyAlaThrSerThr 239  
QY 721 CTGCAGAGAGATCGCCTGATGACCAAGACCCCGCATCCCGTGCGGCGCATCTAC 780  
Db 240 LeuGlnGlnGlnIleAlaTrpMetThrSerAsnProProValProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGCGCTGAACCAAGATCGTGCGGATGACAGCCCGTGAGCATC 840  
Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279  
QY 841 CTGCACATCAACAGAGGCGCCCAAGAGCCCTCCGCGACTAGCTGAGACCGCTTCTCAAG 900  
Db 280 LeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGCAGACAGACACCCAGAGGTGAAGACTGATGACCGCACCCCTGCTG 960  
Db 300 ThrLeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrGluThrLeuLeu 319  
QY 961 GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCGCGCCAGC 1020  
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339  
QY 1021 CTGGAAGAGATGACCGCCTGCGCAAGGCGCGCGCGCCAGCCACAAAGCGCGCTG 1080  
Db 340 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgVal 359  
QY 1081 CTGGCCGAGCGCATGAGCCAGGCCAACACC---AGCGTATGATGCAAGAGCAACTTC 1137  
Db 360 LeuAlaGlnAlaMetSerGlnAlaAsnSerValAsnIleMetGlnLysSerAsnPhe 379  
QY 1138 AAGGGCGCGCGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGCGCCACATCGCCGC 1197  
Db 380 LysGlyProArgArgAsnValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaLys 399  
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGCAGGCAAGAGGCGCCACGATG 1257  
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisGlnMet 419  
QY 1258 AAGCACTGCACCGAGCGCCAGGCCCAACTCTCGGGCAAGATCGGCCCAAGAGGCGC 1317  
Db 420 LysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439  
QY 1318 CGCCCGCGCACTTCTGCAAGACCGCGCGAG-----CCGACCGCC 1359  
Db 440 ArgProGlyAsnPheLeuGlnAsnArgSerGluProAlaAlaProThrValProThrAla 459  
QY 1360 CCCCCCGCGAGAGCTTCCGCTTCAAGAGAGACCAACCCCGCGCCAGAGAGAGCAAG 1419  
Db 460 ProProAlaGlnSerPheArgPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 479  
QY 1420 GAC-----CGGAGACCTGACCAAGCCTGAAGAGCCTGTTGGCAAGACCC 1467  
Db 480 AspArgGlnProTyrArgGlnProLeuThrAlaLeuArgSerLeuPheGlySerGlyPro 499  
QY 1468 CTGAGCCAG 1476  
Db 500 LeuSerGln 502

RESULT 15  
US-10-339-217-126  
; Sequence 126, Application US/10339217  
; Publication No. US20030198648A1

GENERAL INFORMATION:  
APPLICANT: Buechter, Douglas  
APPLICANT: Hou, Xiaohong  
APPLICANT: Marlor, Christopher W.  
APPLICANT: Rice, William G.  
APPLICANT: Yang, Wengang  
TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of  
TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA  
FILE REFERENCE: 111021.143(ACH-US1)  
CURRENT APPLICATION NUMBER: US/10/339,217  
PRIOR FILING DATE: 2003-01-09  
PRIOR APPLICATION NUMBER: US 60/347,369  
NUMBER OF SEQ ID NOS: 144  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 126  
LENGTH: 494  
TYPE: PRT  
ORGANISM: HIV-1  
US-10-339-217-126

Alignment Scores:  
Pred. No.: 1.07e-114 Length: 494  
Score: 2328.50 Matches: 440  
Percent Similarity: 94.56% Conservative: 29  
Best Local Similarity: 88.71% Mismatches: 20  
Query Match: 84.00% Indels: 7  
DB: 14 Gaps: 5  
US-09-475-704A-3 (1-1479) x US-10-339-217-126 (1-494)

QY 4 GCGCGCCCGCCGAGATCCTGCGCGCGCGCAAGCTGGAACGCTGGGAGCGCATCCGCTG 63  
DB 1 GlyAlaArgAlaSerValLeuSerGlyGlyLeuAspAlaTrpGluysileArgLeu 20  
QY 64 CCGCGCGCGCGCAAGTGTCTACATGATGAAGCACTGTGTGGGCCAGCCGCGAGCTG 123  
DB 21 ArgProGlyGlyLysLysLysTyrlsLeuLysHisLeuValTrpAlaSerArgGluLeu 40  
QY 124 GAGAAAGTTCGCTGAAACCCCGCTGCTGAGAGACCGAGGCTGCAAGCAGATCATC 183  
DB 41 GluArgPheAlaLeuAsnProGlyLeuLeuGluThrThrGluGlyCysArgGlnIle 60  
QY 184 CGCGACGTCACCCCGCTGACAGCCGCGAGCGAGCTGAAGAGCTTCAACACC 243  
DB 61 ThrGlnIleGlnProSerIleGlnThrGlySerGluGluIleLysSerLeuTyrsnThr 80  
QY 244 GTGGCCACCTGTACTGCGTGACAGAAAGATCGAGTCCGCGACCAAGAGGCGCTG 303  
DB 81 IleAlaValLeuTyrlsPheValHisGlnLysIleGluValLysAspThrLysGluAlaLeu 100  
QY 304 GACAAAGATCGAGAGAGAGCAGAAAGATGCGAGCAAGATCCAGCAGGCGCGAGCGCC 363  
DB 101 AspLysLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 119  
QY 364 GACAAAGGCAAGGTGAGCCAGAACTACCCCATCGTGCAAGACCTGCAAGGCGCGAGTGTG 423  
DB 120 AspLysGly--ValSerGlnAsnTyrlsProIleValGlnAsnLeuGlnGluMetVal 138  
QY 424 CACCAAGGCGCATCAGCCCGCGACCCCTGAACGCTGCGTGAAGGTGATCGAGAGAGGCC 483  
DB 139 HisGlnAlaLeuSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla 158  
QY 484 TTCAGCCCGGAGGTGATCCCATGTTCACCGCGCTGAGCGAGGCGCCACCCCGCAGGAC 543  
DB 159 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 178  
QY 544 CTGAACACGATGTTGAACACCGTGCGCGCGCCACAGCGCGCGCATGAGTGAAGAC 603  
DB 179 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 198  
QY 604 ACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCCATC 663  
DB 604 ACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCCATC 663

DB 199 ThrIleAsnAspGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle 218  
QY 664 GCGCGCGCGCAGATGCGCGAGCGCGCGCGCGCAGCGACATCGCGCGCACCAAGACCTG 723  
DB 219 ProProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 238  
QY 724 CAGAGCAGATCGCTGGATGACCGCAACCCCGCATCCCGTGGGCGCATCTACAG 783  
DB 239 GlnGluGlnIleAlaTrpMetThrSerAsnProProValProValGlyGluIleTyrls 258  
QY 784 CCGTGGATCATCTGGGCGCTGAACAGATCGTGGCATGTACAGCCCGCTGAGCATCTG 843  
DB 259 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrlsSerProValSerIleLeu 278  
QY 844 GACATCAAGCAGGCGCGCGAGAGCGCTTCCGCGACTACGTGAGCGCTTCTTCAAGACC 903  
DB 279 AspIleArgGlnGlyProLysGluProPheArgAspTyrlsValAspArgPhePheLysThr 298  
QY 904 CTGCGCGCGGAGCAGAGCACCCAGAGGTGAAGACTGTGATGACCGACACCTGTGCTG 963  
DB 299 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuVal 318  
QY 964 CAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCGCGCGCGCGCAGCTG 1023  
DB 319 GlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGlyAlaSerLeu 338  
QY 1024 GAGAGATGATGACCGCTGCGAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCTGCTG 1083  
DB 339 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgIleLeu 358  
QY 1084 GCGGAGCGGATGAGCCAG--GCCAACACACAGCGTGTATGACAGAACCACTTCAAG 1140  
DB 359 AlaGluAlaMetSerGlnValThrAsnProValValMetMetGlnLysGlyAsnPheLys 378  
QY 1141 GCGCGCGCGCGCATCTGCAAGTGTCTCAACTGCGCGCGCAAGAGGCGCACATGCGCGCAAC 1200  
DB 379 GlyHisArgLysIleValLysCysPheAsnLysGlyLysGluGlyHisIleAlaArgAsn 398  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCTGCTGAACTGCGCGCAAGAGGCGCACATGCGCGCAAG 1260  
DB 399 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleGlnMetLys 418  
QY 1261 GACTGACCGAGCGCGCAGGCCAACTTCTGGGCAAGATCTGCGCGCACCAAGGCGCGC 1320  
DB 419 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 438  
QY 1321 CCGGCAACTTCTGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCGC 1380  
DB 439 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheGly 458  
QY 1381 TTC---GAGGAGACCAACCCCGCGCGCAGAGCAGAGAGAGCAAGCAGCGAG----- 1428  
DB 459 PheGlyGluGluIleThrProSerProArgGlnIleThrLysAspLysGluGlnSerPro 478  
QY 1429 ACCCTGACCAAGCTGAAGAGCTGTCGCGCAAGCAGACCCCGTGAAGCAG 1476  
DB 479 ProLeuThrSerLeuLysSerLeuPheGlyAsnAspProLeuSerGln 494

Search completed: March 11, 2005, 16:38:15  
Job time : 200.728 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:05:25 ; Search time 41.8258 Seconds  
(without alignments)  
5279.323 Million cell updates/sec

Title: US-09-475-704A-3  
Perfect score: 2772  
Sequence: 1 atggcgccgcgcgcagcat.....acgacccctgagccagtaa 1479

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US09475704/runat\_10032005\_140222\_14851/app\_query.fasta\_1.3342  
-DB=Issued\_Patents\_AA -QFMT=faстан -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704\_QCGN\_1\_1\_86@runat\_10032005\_140222\_14851 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	2435	87.8	492	4 US-09-991-258-5	Sequence 5, Appli
2	2256	81.4	500	2 US-08-816-155B-45	Sequence 45, Appl
3	2256	81.4	500	3 US-09-079-587-45	Sequence 45, Appl
4	2249	81.1	500	1 US-08-375-510-1	Sequence 1, Appli
5	2249	81.1	500	2 US-08-487-657-1	Sequence 1, Appli
6	2249	81.1	500	4 US-09-309-572-16	Sequence 16, Appl
7	2249	81.1	500	4 US-09-718-096-16	Sequence 16, Appl
8	2245	81.0	1350	4 US-09-952-060-35	Sequence 35, Appl
9	2235	80.6	512	3 US-08-463-210-8	Sequence 8, Appli
10	2235	80.6	512	3 US-09-124-900-2	Sequence 2, Appli
11	2235	80.6	512	4 US-08-463-028-8	Sequence 8, Appli
12	2106	76.0	493	4 US-09-952-060-33	Sequence 33, Appl

13	2037.5	73.5	478	1	US-08-127-499A-11	Sequence 11, Appl
14	2037.5	73.5	478	1	US-08-482-847-11	Sequence 11, Appl
15	2023	73.0	437	3	US-08-392-794A-2	Sequence 2, Appli
16	1976	71.3	512	4	US-09-319-588C-4	Sequence 59, Appl
17	1797.5	64.8	498	1	US-08-470-202-59	Sequence 59, Appl
18	1797.5	64.8	498	1	US-08-471-770-59	Sequence 59, Appl
19	1797.5	64.8	498	2	US-08-468-059-59	Sequence 59, Appl
20	1797.5	64.8	498	3	US-09-109-916-59	Sequence 59, Appl
21	1797.5	64.8	498	4	US-09-886-156-59	Sequence 59, Appl
22	1797.5	64.8	498	4	US-09-886-149-59	Sequence 59, Appl
23	1797.5	64.8	498	4	US-09-886-150-59	Sequence 59, Appl
24	1797.5	64.8	498	4	US-09-886-159-59	Sequence 59, Appl
25	1797.5	64.8	498	4	US-10-326-090-59	Sequence 59, Appl
26	1775.5	64.1	498	1	US-08-470-202-60	Sequence 60, Appl
27	1775.5	64.1	498	1	US-08-471-770-60	Sequence 60, Appl
28	1775.5	64.1	498	2	US-08-468-059-60	Sequence 60, Appl
29	1775.5	64.1	498	3	US-09-109-916-60	Sequence 60, Appl
30	1775.5	64.1	498	4	US-09-886-156-60	Sequence 60, Appl
31	1775.5	64.1	498	4	US-09-886-149-60	Sequence 60, Appl
32	1775.5	64.1	498	4	US-09-886-150-60	Sequence 60, Appl
33	1775.5	64.1	498	4	US-09-886-159-60	Sequence 60, Appl
34	1775.5	64.1	498	4	US-10-326-090-60	Sequence 60, Appl
35	1681	60.6	363	2	US-08-850-049-130	Sequence 130, App
36	1681	60.6	363	3	US-08-050-478-130	Sequence 130, App
37	1681	60.6	363	3	US-09-414-117-130	Sequence 130, App
38	1681	60.6	363	4	US-09-678-437-130	Sequence 130, App
39	1681	60.6	363	4	US-09-943-722-130	Sequence 130, App
40	1536.5	55.4	458	1	US-07-648-796A-5	Sequence 5, Appli
41	1536.5	55.4	799	1	US-07-648-796A-7	Sequence 7, Appli
42	1530.5	55.2	337	1	US-07-648-796A-1	Sequence 1, Appli
43	1514.5	54.6	328	4	US-08-776-188C-77	Sequence 77, Appl
44	1478.5	53.3	518	4	US-09-206-551-45	Sequence 45, Appl
45	1413	51.0	294	3	US-09-370-368-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-09-991-258-5  
Sequence 5, Application US/09991258  
Patent No. 6783939  
GENERAL INFORMATION:  
APPLICANT: Olmsted, Robert  
APPLICANT: Keith, Paula  
APPLICANT: Dryga, Sergey  
APPLICANT: Caley, Ian  
APPLICANT: Maughan, Maureen  
APPLICANT: Johnston, Robert  
APPLICANT: Davis, Nancy  
APPLICANT: Swastrom, Ronald  
TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE  
FILE REFERENCE: 0113.0001U3  
CURRENT APPLICATION NUMBER: US/09/991,258  
CURRENT FILING DATE: 2001-11-16  
PRIOR APPLICATION NUMBER: 09/902,537  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: 60/216,995  
PRIOR FILING DATE: 2000-07-07  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 492  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence; No. 6783939e =  
US-09-991-258-5  
Alignment Scores:  
Pred. No.: 2.97e-141 Length: 492  
Score: 2435.00 Matches: 458

Percent Similarity: 96.96% Conservative: 20  
Best Local Similarity: 92.90% Mismatches: 13  
Query Match: 87.84% Indels: 2  
DB: 4 Gaps: 2

US-09-475-704A-3 (1-1479) x US-09-991-258-5 (1-492)

QY 1 ATGGGCGCCCGCCGAGATCCTCGCGCGGCGGAGCTGGACGCCCTGGAGCGCATCCGC 60  
DB 1 MetAlaAlaArgAlaSerIleLeuArgGluGluLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCCCGCGGAGAGAGTGTACTGATGAGACCTGTGTGGCCAGCCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGAGTTCCGCTGAACCCCGGCTGTGAGACCGAGGCGGTGCAAGACATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGluIle 60  
QY 181 ATCCGCGAGCTGCACCCCGCCCTGCAGACCGGCGAGAGCTGAAGACCTGTTCAC 240  
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80  
QY 241 ACCGTGCGCCACCTGTACTGCGTGCAGAGAGATCGAGGTCCGCGACCAAGAGGCGC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGAGACAGATCGAGAGAGAGAGACAAAGTGCAGAGAAAGATCCAGCAGGCGCGAGCC 360  
DB 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGACAGGGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTCTGAGGGCCAGATG 420  
DB 121 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 139  
QY 421 GTGACACAGGCCATCAGCCCCCGGACCCCTGAAGCGCTGGGTGAAGGTGATCGAGAGAG 480  
DB 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159  
QY 481 GCCTTCAGCCCGGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCGCAG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179  
QY 541 GACCTGAACACGATGTTGAACACCGGTGGCGGCGCCAGCGCCCATGCAATGCTGAAG 600  
DB 180 AspLeuAsnThrMetLeuAsnThrValGlyHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGTGCACGCGCGCGCC 660  
DB 200 AspThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisProValHisAlaGlyPro 219  
QY 661 ATCGCCCCCGGCGAGATGCGCGAGCGCGCGGCGGAGCAATCGCGCGCAACCAACGAC 720  
DB 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGACAGAGAGATCGCTGTGATGACCAACCCCGCATCCCGGTGGGCGACATCTAC 780  
DB 240 LeuGlnGluGlnIleAlaIleTrpMetThrSerAsnProProIleProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAAGATCGTGGCGATGTACAGCCCGGTGAGCATC 840  
DB 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279  
QY 841 CTGACATCAAGCAGGCGCCCAAGAGACCTTCCGCGACTACGTGAACCGCTTCTTCAAG 900  
DB 280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGGAGAGAGACCAAGAGAGGTGAAGAACTGGATGACCGACACCTGTGCTG 960  
DB 300 ThrLeuArgAlaGlnGlnAlaThrGlnGlnValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGACAGACGCGCAACCCGACTGCAGAACCATCTCTGCGCGCTCTCGGCGCGCGCGCAGC 1020

DB 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339  
QY 1021 CTGAGAGAGATGATACCGCTGCGCAGGCGGTGGCGGCGCCAGCCACAGGCGCGCTG 1080  
DB 340 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgVal 359  
QY 1081 CTGCGCGAGCGATGAGCCAGGCGCAACACC--AGCTGATGATGCAAGAGCAACTTC 1137  
DB 360 LeuAlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPhe 379  
QY 1138 AAGGCGCGCGCGCGCATCGTCAAGTCTTCACTGCGGCAAGAGAGGCGCACATGCGCCG 1197  
DB 380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGCAGAGAGAGGCGCACAGATG 1257  
DB 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
QY 1258 AAGGACTGCACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCGGCCAGCCAGAGGCG 1317  
DB 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439  
QY 1318 CGCGCGCGCAACTTCTTGCAGAGCGCGCGCGCGCCAGCGCGCGCGCGCGAGACTTC 1377  
DB 440 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProAlaGlnSerPhe 459  
QY 1378 CGCTTCGAGAGACCAACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
DB 460 ArgPheGluGluThrThrProAlaProLysGlnGluProIleGluArgGluProLeuThr 479  
QY 1438 AGCCTGAAGAGCTGTTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476  
DB 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 2

US-08-816-155B-45  
Sequence 45, Application US/08816155B  
Patent No. 5990091  
GENERAL INFORMATION:  
APPLICANT: TARTAGLIA, JAMES  
APPLICANT: COX, WILLIAM I.  
APPLICANT: GETTIG, RUSSELL R.  
APPLICANT: MARTINEZ, HECTOR  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: PINCUS, STEVEN E.  
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
STREET: 745 FIFTH AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10151  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,155B  
FILING DATE: 12-MAR-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: KOWALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2990  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:

; LENGTH: 500 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: n/a  
; TOPOLOGY: linear  
; MOLECULE TYPE: amino acid  
US-08-816-155B-45

## Alignment Scores:

Pred. No.:	2.59e-130	Length:	500
Score:	2256.00	Matches:	424
Percent Similarity:	92.42%	Conservative:	39
Best Local Similarity:	84.63%	Mismatches:	28
Query Match:	81.39%	Indels:	10
DB:	2	Gaps:	5

US-09-475-704A-3 (1-1479) x US-08-816-155B-45 (1-500)

```
QY 1 ATGGGCGCCGCGCGCATCTCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGluLeuAspArgTrpGluLysIleArg 20
QY 61 CTGGCGCCCGCGCGCAAGAGTGTATGATGAAGCACTGTGTGGCCAGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40
QY 121 CTGAGAGAGTTCCGCCGTAACCCCGGCTGTGAGACCAAGGCGTGCAGCAGATC 180
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60
QY 181 ATCCGCGAGCTGCACCCCGCCCTGCAGACCGGCAAGAGAGCTGAAGACCTGTTCAC 240
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCCCTTAATGCGTGCAAGAGATCGAGTCCGCCACACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100
QY 301 CTGCAACAAGATCGAGAGAGAGAGACAAGTCCAGCAAGATCCAGCAGCGCGAGGCC 360
Db 101 LeuAspLysIleGluGluGlnAsnLysSerLysLysAlaGlnGlnAla---Ala 119
QY 361 GCCGACAAAGGCG-----AAGTGAAGCCAGAACTACCCCATCGTGCAAGACCTGCAG 411
Db 120 AlaAspThrGlyHisSerAsnGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139
QY 412 GGGCAGATGTGCAACAGGCCCATCAGCCCCCGACCCCTGAACGCTGGGTGAAGGTGATC 471
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValVal 159
QY 472 GAGGAGAAAGGCTTCAGCCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCC 531
Db 160 GluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179
QY 532 ACCCCCCAGGACCTGAACAGGATGTTGAACACCGTGGGCGGCCACCAAGCGCCCATGCAG 591
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGln 199
QY 592 ATGCTGAAGGACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGCACCCCGTGAC 651
Db 200 MetLeuLysGluThrIleAsnGluGluAlaIleGluTrpAspArgValHisProValHis 219
QY 652 GCGGCGCCCATCGCGCGCGCGCGAGATGCGGAGCGCGCGGCGGAGGACATGCGCGGAC 711
Db 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239
QY 712 ACCAGCACCTGCAAGAGCAGATCGCTGATGACCAACCCCCCATCCCGGTGGGC 771
Db 240 ThrSerThrLeuGlnGluGlnIleGlyTyrMetThrAsnAsnProIleProValGly 259
QY 772 GACATCTACAAGCGGTGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCCCC 831
Db 260 GluIleTyrLysArgGlyTrpIleIleLeuGlyLeuAsnLysIleValAlaArgMetTyrSerPro 279
```

```
QY 832 GTGACATCTGTGACATCAAGCAGGAGGCCCAAGAGCCCTTCCGCACTACGTGAGCCGC 891
Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299
QY 892 TTCTTCAAGACCCCTGCGCGCGCGAGAGAGACCAACCAAGAGGTGAAGAACTGATGACCCAG 951
Db 300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGlu 319
QY 952 ACCCTGCTGTGCAAGAACGCCAACCCCACTGCAAGACCATCTCTGCGCGCTCGGCCCC 1011
Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339
QY 1012 GCGCGCCAGCTGAGAGAGATGATACCGCTGCCAGGCGCGTGGGCGGCCCAAGCAAG 1071
Db 340 AlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359
QY 1072 GCGCGCGTGTGCGCGAGCGGATGAGCCAGGCCAACCAACAGC-----GTGATGATGCAG 1125
Db 360 AlaArgValLeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGln 379
QY 1126 AAGAGCAACTTCAAGGCGCGCGCGCATGCTCAAGTGTCTTCAACTGCGCAAGAGGCG 1185
Db 380 ArgGlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCySGlyLysGluGly 399
QY 1186 CACATGCGCGCAACTGCGCGCGCGCGCGCATGCTCAAGTGTCTTCAACTGCGCAAGAG 1245
Db 400 HisThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlu 419
QY 1246 GGGCAGCAGATGAAGAGACTGCACGAGCGCGCAACTTCTGGGCAAGATCTGGCCC 1305
Db 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439
QY 1306 AGCCACAAGGCGCGCGCGCGCAACTTCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCG 1365
Db 440 SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459
QY 1366 GCCGAGAGCTTCCGC-----TTGAGAGAGACCAACCCCGCGCAAGAGAGAGCAAG 1419
Db 460 GluGluSerPheArgSerGlyValGluThrThrProProGlnLysGlnGluProIle 479
QY 1420 GACCGCGAG-----ACCTGACAGCCTGAAGAGCTGTTGGCAAGAGACCCCTGAGC 1473
Db 480 AspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsnAspProSerSer 499
QY 1474 CAG 1476
Db 500 Gln 500
```

RESULT 3  
US-09-079-587-45  
; Sequence 45, Application US/09079587  
; Patent No. 613006  
; GENERAL INFORMATION:  
; APPLICANT: TARTAGLIA, JAMES  
; APPLICANT: COX, WILLIAM I.  
; APPLICANT: GETTIG, RUSSELL R.  
; APPLICANT: MARTINEZ, HECTOR  
; APPLICANT: PAOLETTI, ENZO  
; APPLICANT: PINCUS, STEVEN E.  
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
; STREET: 745 FIFTH AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10151  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/079,587  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/816,155  
 FILING DATE: 12-MAR-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KOWALSKI, THOMAS J.  
 REGISTRATION NUMBER: 32,147  
 REFERENCE/DOCKET NUMBER: 454310-2990  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-588-0800  
 TELEFAX: 212-588-0500  
 INFORMATION FOR SEQ ID NO: 45:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 500 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: n/a  
 TOPOLOGY: linear  
 MOLECULE TYPE: amino acid  
 US-09-079-587-45

Alignment Scores:  
 Pred. No.: 2.59e-130 Length: 500  
 Score: 2256.00 Matches: 424  
 Percent Similarity: 92.42% Conservative: 39  
 Best Local Similarity: 84.63% Mismatches: 28  
 Query Match: 81.39% Indels: 10  
 DB: 3 Gaps: 5

US-09-475-704A-3 (1-1479) x US-09-079-587-45 (1-500)

QY 1 ATGGGGCGCCGCGCCGACATCTCCGCGCGGCAAGCTGGACGCGCTGGAGCGCATCCGC 60  
 Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGluLeuAspArgTyrGluLysIleArg 20  
 QY 61 CTGCGCGCGCGCGCGCAAGTGTACATGATGAAGCACTGTGTGGGCGCCAGCGCGAG 120  
 Db 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTyrPalaSerArgGlu 40  
 QY 121 CTGAGAAGTTCGCGCTGAACCCCGGCTGTGAGACCAAGCGGCGCTGCAAGCAGATC 180  
 Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
 QY 181 ATCCGCCAGCTGCACCCCGCGCTGCAAGCGGCGAGGAGGTGAAGCCTGTTCAC 240  
 Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80  
 QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGTCCCGCAGACCAAGAGGCC 300  
 Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100  
 QY 301 CTGACAAGATCGAGAGAGAGCAAAAGTCCAGCAAGAGATCCAGCAGCGCGAGGCC 360  
 Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGlnAla--Ala 119  
 QY 361 GCCGACAAGGC-----AAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGCAG 411  
 Db 120 AlaAspThrGlyHisSerAsnGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139  
 QY 412 GCCCAGATGTGACCAAGGCCATCAAGCCCGCAGCCCTGAACGCTGGGTGAAGTATC 471  
 Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysVal 159  
 QY 472 GAGGAGAAGGCTTCAAGCCCGAGGTGATCCCATGTTCACCGCGCTGAGCGAGGCGGCC 531  
 Db 160 GluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179  
 QY 532 ACCCCCCAGACTGAACACGATGTGAACACCGGTGGCGGCGCACAGGCGCGCATGCAG 591  
 Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGln 199

QY 592 ATGCTGAAGACACCATCAACGAGAGCGCGCGAGTGGAGCCGCGTGACCCCGTGAC 651  
 Db 200 MetLeuLysGluThrIleAsnGluGluAlaIleGluTyrPaspArgValHisProValHis 219  
 QY 652 GCCGCGCCCATCGCGCGCGCGCGCAATGCGGAGCGCGCGCGAGCATCGCGCGCAC 711  
 Db 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239  
 QY 712 ACCAGCACCTTGAGAGAGACAGATCGCCTGATGACCAACACCCCGCATCCCGTGCGC 771  
 Db 240 ThrSerThrLeuGlnGluGlnIleGlyTyrMetThrAsnAsnProIleProValGly 259  
 QY 772 GACATCTACAAGCGGTGATCATCTCGGCGCTGAACAAGATCGTGGATGTACAGCCGC 831  
 Db 260 GluIleTyrLysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279  
 QY 832 GTGAGCATCTTGACATCAAGAGAGCGCGCGCAAGAGCGCGCTTCGCGACTAGTGACCGC 891  
 Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299  
 QY 892 TTCTTCAAGACCTTCGCGCGCGCGCAAGAGACCAAGAGGTGAAGAACTGTGATGACCGAC 951  
 Db 300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTyrMetThrGlu 319  
 QY 952 ACCCTGCTGTGTCAGAACCGCCCAACCCGACTGCAAGACCATCTCGCGCGCTTCGCGCC 1011  
 Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339  
 QY 1012 GCGCGCAGCGCTGAGAGAGATGATGACCGCGCTGCGAGGCGGTGGCGCGCGCACCAAG 1071  
 Db 340 AlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359  
 QY 1072 GCGCGCGTGTGCGCGCGAGCGCATGAGCCAGCCCAACACCGC-----GTGATGATGAC 1125  
 Db 360 AlaArgValLeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGln 379  
 QY 1126 AAGAGCACTTCAAGGCGCGCGCGCGCGCATCTCAAGTGTCTCAACTGCGCGCAAGAGGC 1185  
 Db 380 ArgGlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCysGlyLysGluGly 399  
 QY 1186 CACATCGCGCGCAACTGCGCGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGGCGCAAGAG 1245  
 Db 400 HisThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGlu 419  
 QY 1246 GCGCCACCAAGTGAAGAGATGACCGCGCGCGCGCGCAACTTCTTGCGCAAGATCTGCGCC 1305  
 Db 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTyrPro 439  
 QY 1306 AGCCACAAGGCGCGCGCGCGCAACTTCTTGCGCAAGCGCGCGCGCGCGCGCGCGCGCC 1365  
 Db 440 SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459  
 QY 1366 GCCGAGAGCTTCGCG-----TTGAGAGAGACCAACCCCGCGCGCAAGAGAGAGCAAG 1419  
 Db 460 GluGluSerPheArgSerGlyValGluThrThrThrProProGlnLysGlnGluProIle 479  
 QY 1420 GACCGCGAG-----ACCTGACCAAGCTGGAAGAGCCTGTTCGCGCAAGACCCCGTGA 1473  
 Db 480 AspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsnAspProSerSer 499  
 QY 1474 CAG 1476  
 Db 500 Gln 500

RESULT 4  
 US-08-375-510-1  
 Sequence 1, Application US/08375510  
 Patent No. 5576421  
 GENERAL INFORMATION:  
 APPLICANT: Saito, Atsushi  
 APPLICANT: Sinagawa, Hideo  
 APPLICANT: Nakata, Atsuo

```

1      TITLE OF INVENTION:  HIV ANTIGEN
2      NUMBER OF SEQUENCES:  4
3      CORRESPONDENCE ADDRESS:
4      ADDRESSEE:  Birch, Stewart, Kolaesch and Birch
5      STREET:  P.O. Box 747
6      CITY:  Falls Church
7      STATE:  Virginia
8      COUNTRY:  U.S.A.
9      ZIP:  22040-0747
10     COMPUTER READABLE FORM:
11     MEDIUM TYPE:  floppy disk
12     COMPUTER:  IBM PC compatible
13     OPERATING SYSTEM:  PC-DOS/MS-DOS
14     SOFTWARE:  PatentIn Release #1.0, Version #1.2
15     CURRENT APPLICATION DATA:
16     APPLICATION NUMBER:  US/08/375,510
17     FILING DATE:
18     CLASSIFICATION:  435
19     PRIOR APPLICATION DATA:
20     APPLICATION NUMBER:  US/07/985,949
21     FILING DATE:  04-DEC-1992
22     ATTORNEY/AGENT INFORMATION:
23     NAME:  Svensson, Leonard R.
24     REGISTRATION NUMBER:  30,330
25     REFERENCE/DOCKET NUMBER:  216-309P
26     TELECOMMUNICATION INFORMATION:
27     TELEPHONE:  (703) 241-1300
28     TELEFAX:  (703) 241-2848
29     TELEX:  248345
30     INFORMATION FOR SEQ ID NO:  1:
31     SEQUENCE CHARACTERISTICS:
32     LENGTH:  500 amino acids
33     TYPE:  amino acid
34     STRANDEDNESS:  single
35     TOPOLOGY:  linear
36     MOLECULE TYPE:  protein
37     HYPOTHETICAL:  NO
38     ANTI-SENSE:  NO
39     ORIGINAL SOURCE:
40     ORGANISM:  Human immunodeficiency virus type 1
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
```

Db 460 GluGluSerPheArgPheGlyGluGluThrThrProSerGlnLysGlnGluProIle 479  
QY 1420 GACCGCGAG-----ACCTGACCGAGCCTGTTGGCAACGACCCCTGAGC 1473  
Db 480 AspLysGluLeuTyrProLeuAlaSerLeuArgSerLeuPheGlySerAspProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500  
RESULT 5  
US-08-487-657-1  
; Sequence 1, Application US/08487657  
; Patent No. 5834267  
; GENERAL INFORMATION:  
; APPLICANT: Saito, Atsushi  
; APPLICANT: Sinsagawa, Hideo  
; APPLICANT: Nakata, Atsuo  
; TITLE OF INVENTION: HIV ANTIGEN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch and Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,657  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/375,510  
; FILING DATE: 18-JAN-1995  
; APPLICATION NUMBER: US/07/985,949  
; FILING DATE: 04-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svensson, Leonard R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 216-309P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300  
; TELEFAX: (703) 241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; US-08-487-657-1  
Alignment Scores:  
Pred. No.: 6.92e-130 Length: 500  
Score: 2249.00 Matches: 421  
Percent Similarity: 92.22% Conservative: 41  
Best Local Similarity: 84.03% Mismatches: 29  
Query Match: 81.13% Indels: 10  
DB: 2 Gaps: 5  
US-09-475-704a-3 (1-1479) x US-08-487-657-1 (1-500)  
QY 1 ATGGGGCGCCCGCGCAGCATCCCTGCGCGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60

Db ||||| ||||| ::||| ||||| ::||| ||||| ::||| |||||  
1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrpGluLysIleArg 20  
QY CTGGCGCCCGCGGCAAGAGTGTACATGATGAACACCTGTGTGGCCAGCGCGAG 120  
Db ||||| ||||| ::||| ||||| ::||| ||||| ::||| |||||  
21 LeuArgProGlyGlyLysLysGlnTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40  
QY CTGGAGAAGTTCGGCCCTGAACCCCGGCTGTGTGAGACCAAGCGGGCTGCAACAGATC 180  
Db ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGlnGlyCysArgGlnIle 60  
QY ATCCGCCAGCTGCACCCCGCCCTGCAGACCGGCGAGAGCTGAAGAGCTGTCAAC 240  
Db ::||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGluLeuArgSerLeuTyrAsn 80  
QY ACCGTGGCCACCTGTACTGCGGTGCACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300  
Db ||||| ::||| ||||| ::||| ::||| ::||| ::||| ::||| |||||  
81 ThrIleAlaValLeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAla 100  
QY CTGGACAAGATTCGAGGAGGAGGACGAACAAGTGCACGAGAAGATCCAGAGCGCGAGGCC 360  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
101 LeuAspLysIleGlnGluGlnAsnLysSerLysLysAlaGlnGlnAla--Ala 119  
QY GCCGACAAAGGC-----AAGGTGAGCCGAACCTACCCCATCGTGACAACTGCAG 411  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
120 AlaAspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGln 139  
QY GGCAGATGTGTACACGAGGCCATCAGCCCGCGACCTGAACGCTGGGTGAAGTGATC 471  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysVal 159  
QY GAGGAGAAGGCTTCAGCCCGGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCC 531  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
160 GluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179  
QY ACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGCGGCGCACAGCGCCCATGACG 591  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 199  
QY ATGCTGAAGACACCATCAACGAGAGGCGCGGAGTGGGACCGCGTGACCCCGTGAC 651  
Db ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
200 MetLeuLysGluThrIleAsnGlnGluAlaAlaGluTrpAspArgLeuHisProValHis 219  
QY GCCGGCCCATCGCCCCCGCGCAGATGCGCGAGCCCGCGGCGAGCATCGCCGGCAC 711  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
220 AlaGlyProIleAlaProGlyGlnMetArgLupProArgLysSerAspIleAlaGlyThr 239  
QY ACCAGCACCTGCAGAGCAGATCGCCTGATGACAGCAACCCCCCATCCCGTGCGC 771  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
240 ThrSerThrLeuGlnGlnGlnIleGlyTrpMetThrHisAsnProProIleProValGly 259  
QY GACATCTACAAGCGGTGATCATCTCGGCTGAACAAGATCGTGGGATGTACAGCCCC 831  
Db ::||| ||||| ::||| ::||| ::||| ::||| ::||| |||||  
260 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279  
QY GTGAGCATCTGACATCAAGCAGAGGCGCCCAAGAGGCCCTTCGCGACTACGTGACCGC 891  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299  
QY TTCTTCAAGACCTGCGCGCGCGAGCAGACCAAGAGGTGAAGACTGATGACCGCAC 951  
Db ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
300 PheTyrLysThrLeuArgAlaGlnGlnAlaSerGlnGluValLysAsnTrpMetThrGlu 319  
QY ACCCTGCTGTGCAGAACGCCCAACCCCGACTGCAAGACCATCTCGCGGCTCTCGGCC 1011  
Db ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||  
320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339  
QY GCGCGCAGCTTGAGAGATGATGACGCGCTGCGAGGCGGTGGCGGCGCCAGCCACAG 1071  
Db ||||| ::||| ||||| ::||| ::||| ::||| ::||| ::||| |||||  
340 GlyAlaThrLeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359  
QY GCCCGCGTCTGGCGCGGCGATGAGCCAGCCAAC-----ACCAGCGTATGATGACG 1125  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 360 AlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGln 379  
QY 1126 AAGAGCACTTCAGGGGCGCGCATGCTCAAGTGTCACTGCGCAAGAGGGC 1185  
Db 380 LysGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGly 399  
QY 1186 CACATCGCCCGCAACTGCGCGCGCGCGCGCAAGAGGGCTGCTGAAGTGGCGCAAGGAG 1245  
Db 400 HisIleAlaLysAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGlu 419  
QY 1246 GGCCACGATGAGAGACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCC 1305  
Db 420 GlyHisGlnMetLysAspCysThrGlnArgGlnAlaAsnPhelGlnLysIleTrpPro 439  
QY 1306 AGCCACAGGGCGCGCGCGCAACTTCTGTCAGAGCGCGCGCGCGCGCGCGCGCGCGCG 1365  
Db 440 SerHisLysGlyArgProGlyLysAsnPhelGlnSerArgProGluProThrAlaProPro 459  
QY 1366 GCCGAGAGCTTCGCTTC-----GAGGAGACCAACCGCGCGCGCGCGAGAGAGAGCAAG 1419  
Db 460 GluGlnSerPheArgPheGlyGluGlnThrThrThrProSerGlnLysGlnGluProIle 479  
QY 1420 GACCGCGAG-----ACCCTGACCAAGCTGTAAGAGCTGTTCGCAACGACCCCTGAGC 1473  
Db 480 AspLysGluLeuTyrProLeuAlaSerLeuArgSerLeuPheGlySerAspProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500

RESULT 6

US-09-309-572-16  
; Sequence 16, Application US/09309572  
; Patent No. 6440730  
; GENERAL INFORMATION:  
; APPLICANT: Heinrich-Pette-Institut  
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV  
; FILE REFERENCE: P50489  
; CURRENT APPLICATION NUMBER: US/09/309,572  
; EARLIER FILING DATE: 1999-05-11  
; EARLIER FILING DATE: 1998-11-26  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: gag polypeptide  
US-09-309-572-16

Alignment Scores:

Pred. No.: 6.92e-130 Length: 500  
Score: 2249.00 Matches: 421  
Percent Similarity: 92.22% Conservative: 41  
Best Local Similarity: 84.03% Mismatches: 29  
Query Match: 81.13% Indels: 10  
DB: 4 Gaps: 5

US-09-475-704A-3 (1-1479) x US-09-309-572-16 (1-500)

QY 1 ATGGGCGCCCGCGCGCATCTGCGCGCGCGCAAGCTGAGCGCCTGGAGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCCCGCGCGCAAGAGTGTACATGATGAGCACTGTGTGGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysGlnTyrLysLeuLysHisIleValIleAlaSerArgGlu 40  
QY 121 CTGAGAGAAGTTCGCCCTGAACCCCGCGCTGCTGAGACCAAGGAGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60

QY 181 ATCCGCCAGCTGCACCCCGCCCTGACAGCCGCGAGAGAGAGCTGAAGACCTGTTCAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnLysLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCCCTGTAATGCGGTGACGAGAGATGAGGTCCGCGCACCAAGAGAGCC 300  
Db 81 ThrIleAlaValLeuTyrCysValHisGlnArgIleAspValLysPheThrLysGlnAla 100  
QY 301 CTGAGCAAGATCGAGAGAGAGAGCAACAAGTCCGACGAGAGATCCAGCAGCGCGAGCC 360  
Db 101 LeuAspLysIleGlnGlnGlnLysSerLysLysAlaGlnGlnAla---Ala 119  
QY 361 GCCGCAAGAGC-----AAGGTGAGCCAGAACTACCCCATCTGTGCAAGAACTTCAG 411  
Db 120 AlaAspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGln 139  
QY 412 GGCCGATGTGTCACCAAGCCATCAGCCCGCGCACCCCTGAAAGCCTGGGTGAAGTGATC 471  
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValVal 159  
QY 472 GAGGAGAAAGCCTTCAGCCCGCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGCGCC 531  
Db 160 GluGlnLysAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnGlyAla 179  
QY 532 ACCCCCGAGACCTGAACACGATGTTGAACACCGTGGGCGCGCACCGCCCATGAG 591  
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGln 199  
QY 592 ATGCTGAAGAGACACCATCAACGAGAGCGCGCGAGTGGAGCCGCGTGCACCCCGTGCAC 651  
Db 200 MetLeuLysGlnThrIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHisProValHis 219  
QY 652 GCCGCGCCCATCGCGCGCGCGCGAGATGCGGAGCGCGCGCGAGCATCGCGCGACCC 711  
Db 220 AlaGlyProIleAlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGlyThr 239  
QY 712 ACCAGACCTTGACGAGAGCAGATCGCTGATGACGACCAACCCCATCCCGTGGGC 771  
Db 240 ThrSerThrLeuGlnGlnGlnIleGlyTrpMetThrHisAsnProIleProValGly 259  
QY 772 GACATCTCAAGCGGTGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCCCC 831  
Db 260 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValAlaArgMetTyrSerPro 279  
QY 832 GTGAGCATCTTGACATCAAGCAGAGCGCCCAAGAGCCCTTCGCGACTACGTGACCGC 891  
Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArg 299  
QY 892 TTCCTCAAGACCTTCGCGCGCGCGCGAGAGCAGCAGGAGGTGAAGAACTGATGACCGAC 951  
Db 300 PheTyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTrpMetThrGlu 319  
QY 952 ACCGTGCTGTGAGAAAGCCAAACCCGACTGCAAGACCATCTCGCGGCTTCGCGCCC 1011  
Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339  
QY 1012 GGCGCCAGCTTGAGAGAGATGATGACCGCTGCCAGGCGCGTGGGCGCGCCAGCCACAAG 1071  
Db 340 GlyAlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359  
QY 1072 GCCCGCTGTGCGCGAGCGCATGAGCGCAAC-----ACCAGCTGATGATGACAG 1125  
Db 360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnProAlaThrIleMetIleGln 379  
QY 1126 AAGAGCAACTTCAGGGCGCGCGCGCATGCTCAAGTGTTCATGCGCAAGAGAGGC 1185  
Db 380 LysGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGlnGly 399  
QY 1186 CACATCGCCCGCAACTGCGCGCGCGCGCGCAAGAGGGCTGCTGAAGTGGCGCAAGGAG 1245  
Db 400 HisIleAlaLysAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGlu 419

QY 1246 GGCACCATGATGAGACTGCAACCGCGCCAGCCCACTTCTTGCGCAAGATCTGCCC 1305  
Db 420 G|YHISG|INMet|Lys|Asp|Cys|Thr|G|L|u|a|g|I|n|A|a|s|n|P|he|L|e|u|G|L|y|s|I|e|T|p|P|ro 439  
QY 1306 AGCCACAAGGGCCCCCGCGCAACTTCTTGAGAGCCCGCGAGCCCAACCGCCCCCCC 1365  
Db 440 Ser|H|s|L|y|e|G|L|y|a|g|P|ro|G|L|y|L|y|s|I|n|P|he|L|e|u|G|I|n|S|e|r|A|g|P|ro|G|L|u|P|ro|H|a|l|a|P|ro|P|ro 459  
QY 1366 GCCGAGAGCTTCCGCTTC-----GAGGAGACCCACCCCGCGCAAGAGAGAGAGAG 1419  
Db 460 G|L|u|G|I|n|S|e|r|P|he|a|g|P|he|g|L|y|G|L|u|G|I|n|T|H|r|T|H|P|ro|S|e|r|G|I|n|L|y|e|G|I|n|P|ro|I|e 479  
QY 1420 GACCGCGAG-----ACCTTGACCAAGCTTGAGAGCTTGTGCGCAACGACCCCTGAGC 1473  
Db 480 As|P|L|y|e|G|L|e|u|T|y|r|P|ro|L|e|u|a|l|a|S|e|r|L|e|u|a|g|S|e|r|L|e|u|P|he|g|L|y|S|e|r|a|s|P|P|ro|S|e|r|S|e|r 499  
QY 1474 CAG 1476  
Db 500 G|I|n 500

## RESULT 7

US-09-718-096-16  
; Sequence 16, Application US/09718096  
; Patent No. 6589763  
; GENERAL INFORMATION:  
; APPLICANT: Von Laer, Melke-Dorothee  
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV  
; FILE REFERENCE: 35-195  
; CURRENT APPLICATION NUMBER: US/09/718,096  
; CURRENT FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: DE 19856463.5  
; PRIOR FILING DATE: 1998-11-26  
; PRIOR APPLICATION NUMBER: EP 99250415.9  
; PRIOR FILING DATE: 1999-11-25  
; PRIOR APPLICATION NUMBER: US 09/309,572  
; PRIOR FILING DATE: 1999-05-11  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: gag polypeptide  
US-09-718-096-16

## Alignment Scores:

Pred. No.:	6.92e-130	Length:	500
Score:	2249.00	Matches:	421
Percent Similarity:	92.22%	Conservative:	41
Best Local Similarity:	84.03%	Mismatches:	29
Query Match:	81.13%	Indels:	10
DB:	4	Gaps:	5

US-09-475-704A-3 (1-1479) x US-09-718-096-16 (1-500)

QY 1 ATGGGGCCCGCGCGCATCTGCGCGCGCGCGCAAGCTGACCGCTGGAGCGCATCCGC 60  
Db 1 Met|G|L|y|A|a|A|g|A|a|S|e|r|V|a|l|L|e|u|S|e|r|G|L|y|G|L|u|L|e|u|a|s|P|L|y|S|T|p|G|L|u|L|y|S|I|e|a|g 20  
QY 61 CTGGCGCCCGCGCGCAAGAGTGTATCATGATGAGACACCTGTGTGGCGCAGCCGCGAG 120  
Db 21 L|e|u|a|g|P|ro|G|L|y|L|y|S|L|y|S|G|I|n|T|y|r|L|y|S|L|e|u|Y|S|H|I|e|V|a|l|T|p|L|a|S|e|r|A|g|G|L|u 40  
QY 121 CTGAGAGAAGTTCGCCCTGAACCCCGCGCTGTGAGAGACCAAGAGGGCTGCAAGCATC 180  
Db 41 L|e|u|G|I|n|a|g|P|he|a|l|a|V|a|L|a|s|n|P|ro|G|L|y|L|e|u|G|I|n|T|H|S|e|r|G|L|u|G|I|n|C|y|S|a|r|g|G|I|n|I|e 60  
QY 181 ATCCGCGAGCTGACACCCCGCGCTGAGAGACCGCGCAGAGAGAGCTGAAGAGCTGTCAAC 240  
Db 61 L|e|u|G|L|y|G|I|n|L|e|u|G|I|n|P|ro|S|e|r|L|e|u|G|I|n|T|H|G|L|y|S|e|r|G|L|u|G|I|n|L|e|u|a|g|S|e|r|L|e|u|T|y|r|A|s|n 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGACAGAGAATCGAGGTCCCGCAGACCAAGAGAGGCC 300

Db 81 T|H|r|I|e|a|l|a|V|a|l|L|e|u|T|y|r|C|y|S|V|a|l|H|I|S|G|I|n|a|g|I|e|a|S|P|V|a|l|L|y|S|a|S|P|T|H|L|y|S|G|I|u|A|a 100  
QY 301 CTGACAAGATCGAGAGAGAGAGACAGACAAGTGCACGAGAGATCCAGAGGCGGAGGCC 360  
Db 101 L|e|u|a|s|P|L|y|S|I|e|G|L|u|G|I|n|L|y|S|I|n|L|y|S|e|r|L|y|S|L|y|S|a|g|I|n|I|a|a|---A|l|a 119  
QY 361 GCCGACAAAGGC-----AAGGTGAGCCAGAACTACCCCATCTGTGCAAGAACTGCAAG 411  
Db 120 A|l|a|a|s|P|T|H|G|L|y|a|s|n|A|s|n|S|e|r|G|I|n|V|a|l|S|e|r|G|I|n|a|s|n|T|y|r|P|ro|I|e|V|a|l|G|I|n|a|s|n|L|e|u|G|I|n 139  
QY 412 GCGCAGATGTGACCAAGGCCATCAGCCCGCGCAACCTGACCGCTGGGTGAAGTGATC 471  
Db 140 G|L|y|G|I|n|M|e|t|V|a|H|I|S|G|I|n|a|l|a|I|e|S|e|r|P|ro|A|g|T|H|L|e|u|a|s|n|a|l|a|T|p|V|a|l|L|y|S|V|a|l|V|a|l 159  
QY 472 GAGGAGAGGCGCTTCAAGCCCGAGGTGATCCCATGTTACCGCGCTGAGCGAGGCGGCC 531  
Db 160 G|L|u|G|I|n|L|y|S|a|l|a|P|he|S|e|r|P|ro|G|L|u|V|a|l|I|e|P|ro|M|e|t|P|he|S|e|r|a|l|a|L|e|u|S|e|r|G|L|u|G|I|n|a|l 179  
QY 532 ACCCGCCAGAGCTGAAACAGATGTTGAACACCGTGGCGCGCAACGAGCGCGCATGCAG 591  
Db 180 T|H|r|P|ro|G|I|n|a|s|P|L|e|u|a|s|n|T|H|e|L|e|u|a|s|n|T|H|r|V|a|l|G|L|y|G|I|n|H|I|S|G|I|n|a|l|a|I|e|t|G|I|n 199  
QY 592 ATGCTGAAGACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGTGAC 651  
Db 200 M|e|t|L|e|u|L|y|S|G|I|n|T|H|r|I|e|a|s|n|G|L|u|G|I|n|a|l|a|G|I|n|T|p|a|S|P|a|g|L|y|S|e|r|a|s|P|I|e|a|l|a|G|L|y|T|H|r 219  
QY 652 GCGCGCCCATCGCCCCCGCGCAGATGCGCGAGCCCGCGGAGCGACATCGCGCGCAC 711  
Db 220 A|l|a|G|L|y|P|ro|I|e|a|l|a|P|ro|G|L|y|G|I|n|M|e|t|A|r|g|G|I|n|P|ro|A|r|g|L|y|S|e|r|a|s|P|I|e|a|l|a|G|L|y|T|H|r 239  
QY 712 ACCAGCACCTGACAGAGAGATCGCCTGATGACCAAGCAACCCCGCATCCCGTGCGC 771  
Db 240 T|H|r|S|e|r|T|H|L|e|u|G|I|n|G|L|u|G|I|n|I|e|G|L|y|T|p|M|e|t|T|H|r|H|I|a|s|n|P|ro|P|ro|I|e|P|ro|V|a|l|G|L|y 259  
QY 772 GACATCTACAAGCGGTGATCATCTGCGCTGAACAAGATCGTGCAGATGACAGCCC 831  
Db 260 G|L|u|I|e|T|y|r|S|a|g|T|p|L|I|e|I|e|L|e|u|G|L|y|L|e|u|a|s|n|L|y|S|I|e|V|a|l|A|r|g|M|e|t|T|y|r|S|e|r|P|ro 279  
QY 832 GTGAGCATCTGACATCAAGCAGAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGACCGC 891  
Db 280 T|H|r|S|e|r|I|e|L|e|u|a|s|P|I|e|a|g|I|n|G|L|y|P|ro|L|y|S|G|I|n|P|ro|P|he|a|g|S|P|I|y|r|V|a|l|a|S|P|a|g 299  
QY 892 TTCTTCAAGACCTCGCGCGCGCAGACAGACCCAGAGAGTGAAGACTGATGACCGAC 951  
Db 300 P|H|e|T|y|r|L|y|r|T|H|L|e|u|a|g|A|l|a|G|L|u|G|I|n|a|l|a|S|e|r|G|I|n|G|L|u|V|a|l|L|y|S|a|s|n|T|p|M|e|t|T|H|r|G|L|u 319  
QY 952 ACCCTGCTGCGCAGAACGCCAACCCCGCACTGCAAGACCATCTGCGCGCTCTCGGCC 1011  
Db 320 T|H|r|L|e|u|L|e|u|V|a|l|G|I|n|a|s|n|a|l|a|s|n|P|ro|a|S|P|L|y|S|T|H|r|I|e|L|e|u|L|y|S|a|l|a|L|e|u|G|L|y|P|ro 339  
QY 1012 GCGCGCAGCTGAGAGAGATGATGACCGCGCTGCGCAGGCGGTGGCGGCCCAAGCACAAG 1071  
Db 340 G|L|y|A|l|a|T|H|L|e|u|G|L|u|G|I|n|M|e|t|T|H|r|a|l|a|C|y|S|G|I|n|G|L|y|V|a|l|G|L|y|P|ro|G|L|y|H|I|S|L|y|S 359  
QY 1072 GCGCGGTGCGCGCGAGCGATGAGCGCAAGCCAAC-----ACAGCGTATGATGCAG 1125  
Db 360 A|l|a|A|r|g|V|a|l|L|e|u|a|l|a|G|L|u|a|l|a|M|e|t|S|e|r|G|I|n|V|a|l|T|H|r|a|s|n|P|ro|a|l|a|T|H|r|I|e|M|e|t|I|l|e|G|I|n 379  
QY 1126 AAGAGCAACTCAAGGGCGCGCGCGCATGCTCAAGTGTCTCAACTGCGCAAGAGGGC 1185  
Db 380 L|y|S|G|L|y|a|s|n|P|he|a|g|A|s|n|G|I|n|a|g|L|y|S|T|H|r|V|a|l|L|y|S|C|y|S|P|he|a|s|n|C|y|S|G|L|y|S|G|L|u|G|L|y 399  
QY 1186 CACATGCGCGCAACTGCGCGCGCGCGCGCAAGAGGGCTGCTGAAGTGGCGCAAGAG 1245  
Db 400 H|I|S|I|e|a|l|a|L|y|S|a|s|n|C|y|a|r|g|A|l|a|P|ro|a|r|g|L|y|S|L|y|S|G|L|y|S|T|p|L|y|S|C|y|S|G|L|y|S|G|L|u 419  
QY 1246 GGCACACAGATGAAGAGTGCACCGGAGCGCAAGCTTCTTGCGCAAGATCTGCGCC 1305  
Db 420 G|L|y|H|I|S|G|I|n|M|e|t|L|y|S|a|S|P|C|y|S|T|H|G|L|u|a|g|I|n|a|l|a|s|n|P|he|L|e|u|G|L|y|S|I|e|T|p|P|ro 439  
QY 1306 AGCCACAAGGGCGCGCGCGCAACTTCTGACAGAGCGCGCGAGCGCCAGCGCCCCCCC 1365

Db 440 SerHisblysglyArgProglyAsnPhenleuGlnSerArgProGluProThrAlaProPro 459  
QY 1366 GCCGAGAGCTTCCGCTTC-----GAGGAGACCACCCCGCGGCAAGAGAGAGCAAG 1419  
Db 460 GlnGlnSerPheArgPheGlyGlnGluThrThrProSerGlnLysGlnGluProIle 479  
QY 1420 GACCGCGGAG-----ACCCTGACACGAGCTGTTCGGCAACGACCCCTGAGC 1473  
Db 480 AspLysGlnLeuTyrProleuAlaSerLeuArgSerLeuPheGlySerAspProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500

RESULT 8  
US-09-952-060-35  
; Sequence 35, Application US/09952060  
; Patent No. 6733993  
; GENERAL INFORMATION:  
; APPLICANT: Emini, Emilio A.  
; APPLICANT: Youil, Rima  
; APPLICANT: Bett, Andrew J.  
; APPLICANT: Chen, Ling  
; APPLICANT: Kaslow, David C.  
; APPLICANT: Shiver, John W.  
; APPLICANT: Casimiro, Danilo R.  
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV-1-GAG, POL, NEF AND  
; TITLE OF INVENTION: MODIFICATIONS  
; FILE REFERENCE: 20747Y  
; CURRENT APPLICATION NUMBER: US/09/952,060  
; CURRENT FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: PCT/US01/28861  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/317,814  
; PRIOR FILING DATE: 2001-09-07  
; PRIOR APPLICATION NUMBER: 60/279,056  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/233,180  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 1350  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Codon optimized gag-1A pol fusion  
US-09-952-060-35

Alignment Scores:  
Pred. No.: 1.44e-129 Length: 1350  
Score: 2245.00 Matches: 419  
Percent Similarity: 92.00% Conservative: 41  
Best Local Similarity: 83.80% Mismatches: 32  
Query Match: 80.99% Indels: 8  
DB: 4 Gaps: 4

US-09-475-704A-3 (1-1479) x US-09-952-060-35 (1-1350)  
QY 1 ATGGGGCGCCGCGCCAGCATCTGCGGCGGCAAGCTGGAGCGCTGGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlnLeuAspLysTrpGlnLysIleArg 20  
QY 61 CTGGCGCCCGCGCGGCAAGAGTGTATGATGATGAGCACTGTGTGGGCCAGCGCGGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
QY 121 CTGAGAGAGTTGCGCTGGAACCCCGGCTGTGTGAGACGAGGAGGCTGCAAGCAGATC 180  
Db 41 LeuGlnArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60

QY 181 ATCCGCCAGCTGCACCCCGCCCTGCAGACCGGCGAGCGAGAGCTGAAGACCTGTTCAAC 240  
Db 61 LeuGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGAGAAAGATCAGGTCCGCGACCAAGAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnLysIleAspValLysAspThrLysGlnAla 100  
QY 301 CTGACAGATGCAGAGAGAGCAGAACAGTGCACGAGAAAGATCCAGAGAGCGGAGGCC 360  
Db 101 LeuGlnLysIleGlnGlnGlnGlnAsnLysSerLysLysLysAlaGlnGlnAlaAlaAla 120  
QY 361 -----GCCGACAAAGGCGAAGGTGAGCCAGAACTTACCCATCGTGCAAACTGCAGGGC 414  
Db 121 GlyThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140  
QY 415 CAGATGTGCACCAAGCCATCAGCCCGCACCCCTGAACGCTGGTGAAGTGATCGAG 474  
Db 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValAlaGln 160  
QY 475 GAGAGCCCTTCAAGCCCGAGGTGATCCCATGTTACCGCCCTGAGCGAGGCGCCACC 534  
Db 161 GlnLysAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnGlyAlaThr 180  
QY 535 CCCCAGACCTGAACACAGATGTTGAACCGTGCGCGCCACCAAGCGCCATGCAGATG 594  
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMet 200  
QY 595 CTGAAGACACCATCAACGAGAGAGCGCGAGTGGAACCGGTGCACCCCGTGACGCC 654  
Db 201 LeuLysGlnThrIleAsnGlnGlnAlaGlnAlaGlnTrpAspArgLeuHisProValHisAla 220  
QY 655 GAGCCCATCGCCCGCGCCAGATGCGCGAGCGCCCGCGGACGACATCGCGGACCAACC 714  
Db 221 GlyProIleAlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGlyThrThr 240  
QY 715 AGCACCTGCAGAGAGATCGCCTGATGACCAACCCCGCATCCCGTGCGGCGAC 774  
Db 241 SerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsnProProIleProValGlyGln 260  
QY 775 ATCTCAAGCGGTGATCATCTGCGGCTGAACAAGATCGTGCGATGTAAGCCCGCTG 834  
Db 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280  
QY 835 AGCATCCTGACATCAAGCAGAGGCGCCCAAGAGCGCCTTCCGCGACTGATGACCGCTTC 894  
Db 281 SerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhe 300  
QY 895 TTCAAGACCTTGGCGCGCGGAGCAGAGCAACCCAGAGGTGAAGACTGATGACCGCACCC 954  
Db 301 TyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTrpMetThrGlnThr 320  
QY 955 CTGCTGTGCAGAACCCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCCCGGCC 1014  
Db 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340  
QY 1015 GCCAGCTGCAGAGATGATGACCGGCTGCCAGGCGGTGGCGGCCACCAAGAGGCC 1074  
Db 341 AlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlnHisLysAla 360  
QY 1075 CGCGTGTGCGCGGCGGATGAGCCGCAACACAGC-----GTGATGATGCAGAAAG 1128  
Db 361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArg 380  
QY 1129 AGCAACTTCAAGGCGCGCGCGCATCGTCAAGTCTTCACTGCGGCAAGAGGCGCAC 1188  
Db 381 GlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnLysGlyLysValGlyHis 400  
QY 1189 ATCGCCCGCAACTGCGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGGCGCAAGAGGGC 1248  
Db 401 IleAlaLysAsnLysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGly 420  
QY 1249 CACCAATGAAGAGCTGCACCGGAGCGGCGCAACTTCTGCGGCAAGATCTGGCCAGC 1308

Db 421 HisGlnMetLeuYAspCysAsnGluArgGlnAlaAsnPhenLeuGlyLysIleTrpProser 440  
 QY 1309 CACAAGGGCGCCCGGCAACTTCTGACAGCCCGCCCGGACCCCGCCCGCCCGCCCGCC 1368  
 Db 441 HisLysGlyArgProGlyAsnPhenLeuGlnSerArgProGluProThrAlaProProGlu 460  
 QY 1369 GAGAGCTTCGCTTC-----GAGAGAGACACCCCGCCGACAGAGAGAGAGAGAGAC 1422  
 Db 461 GluSerPheArgPheGlyGluGluThrThrProSerGlnLysGlnIleProIleAsp 480  
 QY 1423 CGCGAG-----ACCCTGACCAAGCCTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAG 1476  
 Db 481 LysGluLeuTyrProLeuAlaSerLeuArgSerLeuPheGlyAsnAspProSerSerGln 500

RESULT 9  
 US-08-463-210-8  
 ; Sequence 8, Application us/08463210  
 ; Patent No. 6001977

GENERAL INFORMATION:  
 APPLICANT: CHANG, Nancy T.  
 APPLICANT: GALLO, Robert C.  
 APPLICANT: WONG-STAL, Flossie  
 TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Morgan & Finnegan, L.L.P.  
 STREET: 345 Park Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10154-0053  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/463,210  
 FILING DATE: 05-JUN-1995  
 CLASSIFICATION: 436  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 06/693,866  
 FILING DATE: 23-JAN-1985  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 06/659,339  
 FILING DATE: 10-OCT-1984  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Serunian, Leslie A.  
 REGISTRATION NUMBER: 35,353  
 REFERENCE/DOCKET NUMBER: 2026-4193US2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 758-4800  
 TELEFAX: (212) 751-6849  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 512 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HTLV-III  
 FEATURE:  
 NAME/KEY: Protein  
 LOCATION: 1..512  
 OTHER INFORMATION: /note="gag protein of HTLV-III"  
 US-08-463-210-8

Alignment Scores:  
 Pred. No.: 4.99e-129 Length: 512  
 Score: 2235.00 Matches: 424

Percent Similarity: 90.64%  
 Best Local Similarity: 82.65%  
 Query Match: 80.63%  
 DB: 3  
 Indels: 22  
 Gaps: 7

US-09-475-704A-3 (1-1479) x US-08-463-210-8 (1-512)

QY 1 ATGGGCGCCCGCCGACATCTGCGCGCGGCAAGCTGAGCCGCTGGAGCCGATCCGC 60  
 Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspArgTyrGluIleArg 20  
 QY 61 CTGCGCCCGCGCGCAAGAAGTCTACATGATGAACACCTGGTGTGGCCAGCCGAG 120  
 Db 21 LeuArgProGlyGlyLysLysTyrLysLeuLysHisIleValTyrAlaSerArgGlu 40  
 QY 121 CTGAGAGATTGCGCCTGAACCCCGGCTGCTGAGACCAAGGAGGCTGCAAGCATC 180  
 Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
 QY 181 ATCCGCCAGCTGCACCCCGCCCTGCAAGACCGGCAAGAGAGCTGAAGCCTGTTCAAC 240  
 Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80  
 QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGGTCCGCGACACCAAGAGGCC 300  
 Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluVal 100  
 QY 301 CTGGAACAAGATCGAGAGAGAGACAACAAGTCCAGACAGATCCAGAGCCGAGGCC 360  
 Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysAlaGlnGlnAla--Ala 119  
 QY 361 GCCGACAAGGC-----AAGGTGAGCCCACTACCCCATCTGTGCAGAACTGCAG 411  
 Db 120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139  
 QY 412 GGCCAGATGTGACACCAAGGCCATCAGCCCGCCGACCTGACGCTGGTGAAGTGATC 471  
 Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValVal 159  
 QY 472 GAGGAGAAGGCTTCAGCCCGGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCC 531  
 Db 160 GluGluLysAlaPheSerProGluValIlePrometPheSerAlaLeuSerGluGlyAla 179  
 QY 532 ACCCCCGAGACCTGAACAGATGTTGAACACCGGTGGCGGCCACAGGCCGCGCATGCAG 591  
 Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 199  
 QY 592 ATGCTGAAGACACCATCAAGAGAGAGAGCGCCGAGTGGACCGCGTGCACCCCGTGCAC 651  
 Db 200 MetLeuLysGluThrIleAsnGluGluAlaIleGluTyrAspArgValHisProValHis 219  
 QY 652 GCCGGCCCATCGCCCGCCGACATGCGCGAGAGCCCGCGGACAGCATCGCCGCGCACC 711  
 Db 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239  
 QY 712 ACCAGACCCCTGACAGAGAGATGCGCTGATGACCAAGACACCCCGCATCCCGTGGGC 771  
 Db 240 ThrSerThrLeuGlnGluGlnIleGlyTyrMetThrAsnAsnProProIleProValGly 259  
 QY 772 GACATCTCAAGCGGTGATCATCTGGGCTGAACAAGATCGTGGATGTACAGCCCC 831  
 Db 260 GluIleTyrLysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279  
 QY 832 GTGAGCATCTGACATCAAGCAGAGGCGCCCAAGAGCCCTTCGCGCACTAGTGAGCCGC 891  
 Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299  
 QY 892 TTCTTCAAGACCTGCGCGCCGAGCAGACCAAGAGGTGAAGACTGATGACCGCAC 951  
 Db 300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnIleValLysAsnTyrMetThrGlu 319  
 QY 952 ACCCTGTGTGTGAGAACGCCAACCCGCACTGAAGACCATCTGCGGCTCTCGGCCCC 1011

Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339  
QY 1012 GGGCCAGCCTTGAGAGATGATGACCGCTGCGAGGGCGGTGGCGCCCAAGCACAAG 1071  
Db 340 AlaAlaThrLeuGlnGluMetMetThrAlaCysGlnGlyValGlyProGlyHisLys 359  
QY 1072 GCGCGCTGCTGGCCGAGCGATGAGCCAG---GCCAACAAC---AGCGTATGATGCAG 1125  
Db 360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln 379  
QY 1126 AAGAGCACTTCAAGGGCGCGCGCATGCTCAAGTGTCTCACTGCGGCAAGAGGGC 1185  
Db 380 ArgGlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnCysGlyLysGluGly 399  
QY 1186 CACATCGCCCGCAACTGCGCGCGCGCGCGCAAGAGGGCTGTGAAGTGGCGCAAGAG 1245  
Db 400 HisThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlu 419  
QY 1246 GGGCACCAGATGAAGAGCTGCACCGAGCGCGCAAGGCAACTTCTGGGGCAAGATCTGGCCC 1305  
Db 420 GlyHisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439  
QY 1306 AGCCACAGGGCGCGCGCGCGCAAC----- 1329  
Db 440 SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459  
QY 1330 TTCTGTCAGAGCGCGCGCGAGCGCCACCGCGCGCGCGAGAGCTTCCGC-----TTC 1383  
Db 460 PheLeuGlnSerArgProGluProThrAlaProProGluGlnSerPheArgSerGlyVal 479  
QY 1384 GAGAGACCACCGCGCGCGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
Db 480 GluThrThrThrProProGlnLysGlnGluProIleAspLysGluLeuTyrProLeuThr 499  
QY 1438 AGCTGAAGAGCCTGTTGGCAACGAGACCGCTGAGCCAG 1476  
Db 500 SerLeuArgSerLeuPheGlyAsnAspProSerSerGln 512

RESULT 10  
US-09-124-900-2  
Sequence 2, Application US/09124900  
Patent No. 6268484  
GENERAL INFORMATION:  
APPLICANT: KATINGER, Hermann  
APPLICANT: BUCHACHER, Andrea  
APPLICANT: ERNST, Wolfgang  
APPLICANT: BALAUN, Claudia  
APPLICANT: PURTSCHER, Martin  
APPLICANT: TRKOLA, Alexandra  
APPLICANT: FREDL, Renate  
APPLICANT: SCHMATZ, Christine  
APPLICANT: KLIMA, Annelies  
APPLICANT: STEINDL, Franz  
APPLICANT: MUSTER, Thomas  
TITLE OF INVENTION: HIV-Vaccines  
FILE REFERENCE: 1939-112P  
CURRENT APPLICATION NUMBER: US/09/124, 900  
CURRENT FILING DATE: 1998-07-30  
PRIOR APPLICATION NUMBER: PCT/EP95/01481  
PRIOR FILING DATE: 1995-04-19  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 2  
LENGTH: 512  
TYPE: PRT  
ORGANISM: Human Immunodeficiency virus type 1  
US-09-124-900-2

Alignment Scores:  
Pred. No.: 4.99e-129 Length: 512  
Score: 2235.00 Matches: 424  
Percent Similarity: 90.64% Conservative: 41  
Best Local Similarity: 82.65% Mismatches: 26

Query Match: 80.63% Indels: 22  
DB: 3 Gaps: 7  
US-09-475-704A-3 (1-1479) x US-09-124-900-2 (1-512)  
QY 1 ATGGCGCGCGCGCGAGCATCTTCGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGCAAGAAGTGCTACATGATGAACCACTGTGTGGCGCGAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAAGATTGCGCCTGAACCCCGCGCTGCTGAGACCGAGAGAGAGAGAGAGAGAGAGATC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
QY 181 ATCCGCGAGCTGCACCCCGCGCGCTGAGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCAGCCTGTACTGCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100  
QY 301 CTGCAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysAlaGlnGlnAla---Ala 119  
QY 361 GCCGCAAGGGC-----AAGGTGAGCCAGAGACTACCCCATCTGTCAGAACCTGCAG 411  
Db 120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139  
QY 412 GCGCAGATGTGTACCAAGCCATCAAGCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAG 471  
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValVal 159  
QY 472 GAGGAGAAAGCCTTACAGCCCGCGAGGTGATCCCATGTTCACCGCGCTGAGCGAGCGCC 531  
Db 160 GluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179  
QY 532 ACCCGCCAGGACCTGAACAGATGTGAACACCGTGGCGCGCGCCACAGCGCGCATGCAG 591  
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGln 199  
QY 592 ATGCTGAAGAGACCATCAACGAGAGAGCGCGCGAGTGGAGCCGGGTGCACCCCGTGCAC 651  
Db 200 MetLeuLysGluThrIleAsnGluAlaAlaGluTrpAspArgValHisProValHis 219  
QY 652 GCGCGCGCGCATCGCGCGCGCGCGAGTGCAGAGCGCGCGCGCGAGAGAGAGAGAGAGAG 711  
Db 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239  
QY 712 ACCAGCACCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771  
Db 240 ThrSerThrLeuGlnGluGlnIleGlyTyrMetThrAsnAsnProIleProValGly 259  
QY 772 GACATCTCAAGCGGTGATCATCTGCGCTGAACAAGATCGTGGAGATGTACAGCGCC 831  
Db 260 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279  
QY 832 GTGAGCATCTGAGCATCAAGCAGAGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299  
QY 892 TTCTCAAGACCTTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951  
Db 300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGlu 319  
QY 952 ACCGTGCTGTGAGAAAGCGCAAGCCCGAGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339

```
QY 1012 GCGCCAGCTGAGAGATGATGACCGCTGCGAGGCGGTGGCGCCCGACACAG 1071
Db 340 AlaIatThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGluHisLys 359
QY 1072 GCGCGCTGCTGGCCGAGGCGATGAGCCAG---GCCAACAC---AGCGTATGATGAG 1125
Db 360 AlaArgValLeuAlaGluAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln 379
QY 1126 AAGAGCACTTCAAGGGCGCCCGCGCATGTCAGTGTCTCACTGCGGCAAGAGGCG 1185
Db 380 ArgGlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnCysGlyLysGlu 399
QY 1186 CACATCGCCCGCAACTGCGCGCGCCCGCGCAAGAGGCTGCTGAAGTGGCGCAAGAG 1245
Db 400 HisThrAlaArgAsnCyArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlu 419
QY 1246 GCGCACGATGAAGAGTGCACCGGAGCGCGAGCGCCCACTTCTGGGCAAGATCTGGCC 1305
Db 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439
QY 1306 AGCCACAAGGGCGCGCGCGCAAC----- 1329
Db 440 SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459
QY 1330 TTCCTGAGAGCGCGCGCGAGCGCCACCGCGCGCGCGCGAGAGCTTCCGC-----TTC 1383
Db 460 PheLeuGlnSerArgProGluProThrAlaProProGluGlnSerPheArgSerGlyVal 479
QY 1384 GAGGAGACCAACCCCGCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1437
Db 480 GluThrThrThrProProGlnLysGlnGluProIleAspLysGluLeuTyrProLeuThr 499
QY 1438 AGCCTGAAGAGCCTGTTGGCAACGAGCCCTGAGCCAG 1476
Db 500 SerLeuArgSerLeuPheGlyAsnAspProSerSerGln 512

RESULT 11
US-08-463-028-8
; Sequence 8, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STALL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Plunegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,028
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
; FILING DATE: 23-JAN-1985
; APPLICATION DATA:
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..512
; OTHER INFORMATION: /note= "gag protein of HTLV-III"
US-08-463-028-8

Alignment Scores:
Pred. No.: 4,99e-129 Length: 512
Score: 2235.00 Matches: 424
Percent Similarity: 90.64% Conservative: 41
Best Local Similarity: 82.65% Mismatches: 26
Query Match: 80.63% Indels: 22
DB: 4 Gaps: 7

US-09-475-704A-3 (1-1479) x US-08-463-028-8 (1-512)
QY 1 ATGGGCGCGCGCGCGAGCATCTGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspArgTrpGluLysIleArg 20
QY 61 CTGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGACACCTGTGTGGCGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40
QY 121 CTGAGAGAGTTCGCCCTGAACCCCGCGCTGCGAGAGCGAGGAGGCTGCAAGCAGATC 180
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60
QY 181 ATCCGCCAGCTGCACCCCGCGCTGCAGACCGCGAGGAGGAGCTGAAGCCTGTTCAAC 240
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGACCACTTGTACTGCGTGCACAGAAAGATCGAGGTCCGCGACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100
QY 301 CTGACAAAGATCGAGAGGAGGAGCAACAAGTCCAGCAGAAATCCAGAGCGCGAGGCC 360
Db 101 LeuAspLysIleGluGluGlnAsnLysSerLysLysLysAlaGlnAla---Ala 119
QY 361 GCCGCAAGGCGC-----AAGGTGAGCGCAAGTACCCTACCTGCGAGAACCTGCAG 411
Db 120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139
QY 412 GCGCAGATGTGCACCGAGCATCAGCCCGCGACCCCTGAAGCGCTGAGGTGAGGTGATC 471
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleTrpValLysVal 159
QY 472 GAGGAGAGGCTTCAAGCCCGAGGTGATCCCATGTTCAACCGCGCTGAGCGAGCGCC 531
Db 160 GluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179
QY 532 ACCCGCGAGGAGCTGAACAGATGTTGAACCGTGGCGCGCGACCGCGCATGAG 591
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 199
QY 592 ATGCTGAAGGACACCATCAACGAGAGCGCGCGAGTGGAGCGCGCTGACCCCGTGAC 651
Db 200 MetLeuLysGluThrIleAsnGluGlnAlaIleGluTrpAspArgValHisProValHis 219
QY 652 GCGGCGCGCATCGCCCGCGCGAGATGCGGAGCGCGCGCGAGCGAGCATCGCGGAGCC 711
```

```
Db      220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239
QY      712 ACCAGCACCCTGCAGAGCAGATCGCTGATGACCAACCCCGCCCTCGTGCGC 771
Db      240 ThrSerThrLeuGlnGlnIleGlyTrpMetThrAsnAsnProProIleProValGly 259
QY      772 GACATCTACAAGCGGTGATCATCTGCGCTGAACAAGATCGTGGATGTACAGCCCC 831
Db      260 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279
QY      832 GTGAGCATCTGACATCAAGCAGGCGCCCAAGAGACCCCTCCGCACTACGTGACCGC 891
Db      280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299
QY      892 TTCTCAAGACCTGCGCGCCGAGCAGACCAACCAGAGGTGAAGACTGATGACCGAC 951
Db      300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGlu 319
QY      952 ACCCTGCTGTGCAGAACGCCCAACCCCGACTGCAGACCATCTGCGCTCTCGCCCC 1011
Db      320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339
QY      1012 GCGCGCAGCCTGAGAGATGATGACCGCTGCGCGCTGCGCGCGCCAGCCACAG 1071
Db      340 AlaAlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359
QY      1072 GCGCGCTGTGCTGCGCGCGCGATGAGCCAG--GCCAACACC--AGCGTATGATGACG 1125
Db      360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln 379
QY      1126 AAGAGCAACTTCAAGGGGCGCGCGCATCGTCAAGTCTTCACTGCGCGCAAGAGGCG 1185
Db      380 ArgGlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnGlyLysGlnGly 399
QY      1186 CACATCGCGCCGCAACTGCGCGCGCGCGCGCGCAAGAAGGCTGCTGAAGTGGCGCAAGAG 1245
Db      400 HisThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlu 419
QY      1246 GCGCACCATGATGAAGACTGCACCGCAGCGCCAGGCCCACTTCTGCGCAAGATCTGCGCC 1305
Db      420 GlyHisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439
QY      1306 AGCCACAAGGGCGCGCGCGCGCAAC----- 1329
Db      440 SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459
QY      1330 TTCTGTCAGAGCGCGCGCGAGCCCGCGCGCGCGCGCGCGAGAGCTTCCGC-----TTC 1383
Db      460 PheLeuGlnSerArgProGluProThrAlaProProGlnGlnLysPheArgSerGlyVal 479
QY      1384 GAGGAGACCAACCCCGCGCGCGAGAGAGAGCAAGAGACCGCGAG-----ACCTGACCC 1437
Db      480 GluThrThrThrProProGlnLysGlnGluProIleAspLysGluLeuTyrProLeuThr 499
QY      1438 AGCTGAAGAGCTGTTGGCAACGACCGCCCTGAGCCAG 1476
Db      500 SerLeuArgSerLeuPheGlyAsnAspProSerSerGln 512

RESULT 12
US-09-952-060-33
; Sequence 33, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emml, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Betc, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
```

```
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: tPA-gag fusion open reading frame
US-09-952-060-33
```

```
Alignment Scores:
Pred. No.: 3,79e-121 Length: 493
Score: 2106.00 Matches: 394
Percent Similarity: 92.29% Conservative: 37
Best Local Similarity: 84.37% Mismatches: 28
Query Match: 75.97% Indels: 8
DB: 4 Gaps: 4
```

US-09-475-704A-3 (1-1479) x US-09-952-060-33 (1-493)

```
QY      100 CTGCTGTGGGCGCCAGCGCGGAGCTGAGAGAAGTTCGCCCTGAACCCCGCCCTGTCGAGACC 159
Db      27 TleValTrpAlaSerArgGluLeuGluArgPheAlaValAsnProGlyLeuLeuGluThr 46
QY      160 AGCGAGGCTGCAGCAGATCATCCGCCAGCTGCACCCCGCCCTGACAGACCGGACGAG 219
Db      47 SerGluGlyCysArgGlnIleLeuGlnLeuGlnProSerLeuGlnThrGlySerGlu 66
QY      220 GAGCTGAAGAGCTGTTCACACCCGTGGCCACCCCTGTACTGCTGACGAGAGATCGAG 279
Db      67 GluLeuArgSerLeuTyrAsnThrValAlaThrLeuTyrCysValHisGlnLysIleAsp 86
QY      280 GTCCCGCACACCAAGAGAGGCGCTGACCAAGATCGAGAGAGAGCAGACAAGTGCCAGCAG 339
Db      87 ValLysAspThrLysGlnAlaLeuGlnLysIleGlnGlnGlnAsnLysSerLysLys 106
QY      340 AAGATCCAGCAGCGCGGAGGCC-----GCCGACAAGGGCAAGGTGAGCCAGAACTAATCCC 393
Db      107 LysAlaGlnGlnAlaAlaAlaGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrPro 126
QY      394 ATCTGCAGAACTCTGACGGCGCAGATGTGTGACCAAGCCATCAGCCCGCGACCTGGAAC 453
Db      127 IleValGlnAsnLeuGlnGlnGlnMetValHisGlnAlaIleSerProArgThrLeuAsn 146
QY      454 GCCTGGCTGAAGGTGATCGAGAGAGAGCGCTTACGCCCGCGAGGTGATCCCATGTTACCC 513
Db      147 AlaTrpValLysValValAlaGluGluLysAlaPheSerProGluValIleProMetPheSer 166
QY      514 GCGCTGAGCGAGGCGCCACCCCGCGAGACCTGAACAAGATGTTGAACACCGTGGCGGCG 573
Db      167 AlaLeuSerGlnGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGly 186
QY      574 CACGAGCGCGCATGTCAGATGCTGAAGACCAACCATCAAGAGAGGCGCGCGAGTGGAGC 633
Db      187 HisGlnAlaAlaMetGlnMetLeuLysGluThrIleAsnGlnGluAlaAlaGluTrpAsp 206
QY      634 CGCGTGACCCCGCTGCACGCGCGCGCCATCGCCCGCGCGAGATGCGCGAGCGCGCGCG 693
Db      207 ArgLeuHisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGly 226
```

QY 694 AGCGACATCGCCGCGCACACGACCCCTGCGAGAGAGATCGCTGATGACCAAGC 753  
Db 227 SeraspIlealaglyThrThrSerThrLeuGlnGlnIleGlyTrpMetThrAsnAsn 246  
QY 754 CCCCCCATCCCGTGGCGCATCTACAAGCGGTGATCATCTGGGCTGAACAAGATC 813  
Db 247 ProProIleProValGlyIleTyrIleTyrIleTyrIleLeuGlyLeuAsnIle 266  
QY 814 GTGCGATGTAACAGCCCGCTGAGCATCTGACATCAAGCAGGCGCCCAAGAGCCCTTC 873  
Db 267 ValArgMetTyrSerProThrSerIleLeuAspIleArgGlnGlyProIleProIle 286  
QY 874 CGCGACTACGTGACCGCTTCTTCAAGACCTCGCGCGCGAGAGAGACCAAGAGATG 933  
Db 287 ArgAspTyrValAspArgPheTyrIleThrLeuArgAlaGlnIleAspGlnIleVal 306  
QY 934 AAGAACTGATGACCGCACACCTGCTGTGCGAAGCGCAACCCGACTGCAAGACCATC 993  
Db 307 LysAsnTrpMetThrGlnThrLeuLeuValGlnAsnAlaAsnProAspCysIleThrIle 326  
QY 994 CTGCGCGCTCTCGCGCGCGCGCGCGAGAGATGATGACCGCTGCGAGCGGTG 1053  
Db 327 LeuLysAlaLeuGlyProAlaAlaThrLeuGlnGlnMetMetThrAlaCysGlnIleVal 346  
QY 1054 GCGGCGCGCGCACACAGCGCGCGCTGTGCGCGAGCGATGACGCAACACCAAGC 1113  
Db 347 GlyGlyProGlyIleLysAlaAlaArgValLeuAlaGlnIleAlaMetSerGlnValThrAsnSer 366  
QY 1114 -----GTGATGATGACAGAGACCACTTCAAGGCGCGCGCGCATGTCAGTGTCTC 1167  
Db 367 AlaThrIleMetMetGlnArgGlyAsnArgPheArgAsnGlnArgIleThrValLysCysPhe 386  
QY 1168 AACTGCGCGCAAGAGGCGCACATCGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCTGC 1227  
Db 387 AsnCysGlyLysValGlyIleIleAlaLysAsnCysArgAlaProArgLysLysCys 406  
QY 1228 TGAAGTGGCGCAAGAGGCGCGCACATGAAGACTGCAACGCGCGCGCAAGCTTC 1287  
Db 407 TrpLysCysGlyLysGlnGlyIleIleGlnMetLysAspCysAsnGlnArgGlnAlaAsnPhe 426  
QY 1288 CTGGGCAAGATCTGGCGCGCGCACAGAGGCGCGCGCGCACTTCTGCGAGCGCGCCC 1347  
Db 427 LeuGlyLysIleTrpProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgPro 446  
QY 1348 GAGCCCAACCGCGCGCGCGCGCGAGCTTCCGCTTC-----GAGAGACCAACCGCGCGC 1401  
Db 447 GluProThrAlaProProGlnGlnIleSerPheArgPheGlyGlnIleLysThrThrProSer 466  
QY 1402 CAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1455  
Db 467 GlnLysGlnGlnIleProIleAspLysGlnLeuTyrProLeuAlaSerLeuArgSerLeuPhe 486  
QY 1456 GGCAACGACCCCGCTGAGCCAG 1476  
Db 487 GlyAsnAspProSerSerGln 493

## RESULT 13

US-08-127-499A-11  
Sequence 11, Application US/08127499A  
Patent No. 5510264  
GENERAL INFORMATION:  
APPLICANT: VAN ALSTYNE, Diane  
APPLICANT: SHARMA, Lawrence Rajendra  
TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED  
TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/127, 499A  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 51916/102/INBI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 478 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
US-08-127-499A-11

Alignment Scores:  
Pred. No.: 5,78e-117 Length: 478  
Score: 2037.50 Matches: 388  
Percent Similarity: 90.81% Conservative: 47  
Best Local Similarity: 81.00% Mismatches: 33  
Query Match: 73.50% Indels: 11  
DB: 1 Gaps: 8

US-09-475-704A-3 (1-1479) x US-08-127-499A-11 (1-478)

QY 1 ATGGGCGCGCGCGCGCACATCTGCGCGCGCAAGCTGACCGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspArgTrpGlnLysIleArg 20  
QY 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACCTGTGTGGCGAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
QY 121 CTGAGAAAGTTCGCCCTGAACCCCGCGCTGCTGAGAGACGAGCGAGGCTGCAAGCAGATC 180  
Db 41 LeuGlnArgPheAlaValAsnProGlyLeuLeuGlnIleThrSerGlnGlyCysArgGlnIle 60  
QY 181 ATCCGCGAGCTGCACCCCGCGCTGACAGACCGCGAGAGAGAGAGAGAGAGAGAG 240  
Db 61 LeuGlnGlnLeuGlnProSerLeuGlnIleThrGlySerGlnGlnLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAGAGATGAGCTCCGCGACCAAGAGAGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGlnIleLysAspThrLysGlnIle 100  
QY 301 CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnIleGlnAsnLysSerLysLysLysAlaGlnGlnAla--Ala 119  
QY 361 GCCGACAAGGC-----AAGGTGAGCCAGAACTAACCCCATCGTGACAGAACTGAG 411  
Db 120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139  
QY 412 GCGCAGATGTGACACGAGCCATCAGCCCGCGACCTGAAAGCTGGTGAAGGTGATC 471  
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValVal 159  
QY 472 GAGGAGAGGCTTACGCGCGCGAGGTATCCCATGTTCACCGCTGAGCGAGGCGCC 531  
Db 160 GluGlnLysAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnGlyAla 179  
QY 532 ACCCGCGAGGACCTGAACAGATGTGAACACCGTGGCGCGCGCACGAGCGCGCATGCGAG 591  
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaIleMetGln 199

QY	592	ATGCTGAAGGACACCATCAACGAGAGGGCCGGCGAGTGGAGCCGGCTGCACCCCGCTGCAC	651
		:::	
Db	200	MetLeuLysGluThrIleAsnGluGluAlaIleGluTrpAspArgValHisProValHis	219
		:::	
QY	652	GCCGGCCCCATCGCCCCCGGCCAGATGCGCGAGCCCCGGCGGACGACATCGCCGGCAC	711
		:::	
Db	220	AlaGlyProIleAlaProGlyMetArgGluProArgGlySerAspIleAlaGlyThr	239
		:::	
QY	712	ACCAGCACCCCTGCAGAGCAGATCGCTGGATGATGACCAACCCCCCATCCCGTGGC	771
		:::	
Db	240	ThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnProProIleProValGly	259
		:::	
QY	772	GACATCTACAAGCGGTGATCATCTGGGCTTGAACAAGATCGTCGGATGTACAGCCCC	831
		:::	
Db	260	GluIleTyrLysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro	279
		:::	
QY	832	GTGAGCATCCTGGAATCAAGCAGAGGCCCCAAGAGCCCTTCGCGACTACGTGACCGC	891
		:::	
Db	280	ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg	299
		:::	
QY	892	TTCCTCAAGACCTTCGCGCGCCGACAGACACCCAGAGGATGAACTGATGATGACCGAC	951
		:::	
Db	300	PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnValLysAsnTrpMetThrGlu	319
		:::	
QY	952	ACCCCTGTGTGAGAAAGCCCAACCCGACTGCAAGACCATCTTCGCGCTTCGCGCCC	1011
		:::	
Db	320	ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro	339
		:::	
QY	1012	GCGCGCAGCCTGAGAGAGATGATGACCGCTGCCAGGGCGTGGCGGGCCCCAGCCAAG	1071
		:::	
Db	340	AlaAlaThrLeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys	359
		:::	
QY	1072	GCCCGCGTGTGGCCGAGGCGATGAGCCAG--GCCAACACC--AGCGTATGATGCAG	1125
		:::	
Db	360	AlaArgValLeuAlaGluAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln	379
		:::	
QY	1126	AAGAGCAACTTCAAGGGCCCCCGCGCATCGTCAAGTGCTTCAACTGCGGCAAGAGGC	1185
		:::	
Db	380	ArgGlyAsnPheArgAsnGlnAlaGlyLysMetValLysCysPheAsnCysGlyLysGluGly	399
		:::	
QY	1186	CACATCGCCCCGCACTGCGCGGCCCCCGCAAGAAGGGCTGCTGGAAGTGGCGCAAGAG	1245
		:::	
Db	400	HisThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlu	419
		:::	
QY	1246	GCGCACCAAGATGAAGACTGCACCGAGCGGCCAGGCCAACTTCTGGGCAAGATC--TGG	1302
		:::	
Db	420	GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleCysLeu	439
		:::	
QY	1303	CCCAAGCCACAAGGCGCGCCCCGCGCACTTCTTCGAGAAGCGCGCCCCAGCCACCGCCCC	1362
		:::	
Db	440	ProThrArgGluGlnGln--GlyIlePhePhe-----ArgAlaAspGlnSerGlnGln	456
		:::	
QY	1363	CCCGCCGAGAGCTTCGCTTCAGAGAGACC--ACCCCGCGGCAAGAGCAGAGAGC	1416
		:::	
Db	457	ProHisHisPhePheArgAlaAspGlnSerGlnGlnProHisGlnLysArgAlaSer	475
		:::	

RESULT 14  
 US-08-482-847-11  
 ; Sequence 11, Application US/08482847  
 ; Patent No. 5556757  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT: VAN ALSTYNE, Diane  
 ;  
 ; APPLICANT: SHARMA, Lawrence Rajendra  
 ;  
 ; TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR  
 ;  
 ; TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR  
 ;  
 ; TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF  
 ;  
 ; NUMBER OF SEQUENCES: 40  
 ;  
 ; CORRESPONDENCE ADDRESS:  
 ;  
 ; ADDRESSEE: Foley & Lardner  
 ;  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ;  
 ; CITY: Washington  
 ;  
 ; STATE: D.C.

```

: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/482,847
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/127,499
: FILING DATE: 28-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 51916/104/INBI
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 672-5300
: TELEFAX: (202) 672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 478 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: US-08-482-847-11

Alignment Scores:
Pred. No.: 5,78e-117 Length: 478
Score: 2037.50 Matches: 388
Percent Similarity: 90.81% Conservative: 47
Best Local Similarity: 81.00% Mismatches: 33
Query Match: 73.50% Indels: 11
DB: 1 Gaps: 8

US-09-475-704A-3 (1-1479) x US-08-482-847-11 (1-478)

QY 1 ATGGGCGCCCGCCGACATCTGCGCGCGGCGGACGCTGGACGCTGGGAGCGCATCCGC 60
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTyrGlyLysIleArg 20
QY 61 CTGGCGCCCGCGCGGACAGAGTGTACATGATGAAGCAACCTGTGTGGCCAGCCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTyrAlaSerArgGlu 40
QY 121 CTGGAGAAAGTTGCGCCCTGAACCCCGGCGCTGTGGAGACCAGCGAGGCTGCAGACAGATC 180
Db 41 LeuGluArgPheAlaValAlaAsnProGlyLeuLeuGluThrSerGluThrCysArgGlnIle 60
QY 181 ATCCGCCAGCTGCACCCCGCCCTGCAGACCCGCGACGAGGAGAGTGAAGAGCTGTTCAAC 240
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGTCGTGCAGCAGAAAGATCGAGGTCCGCGACACCAAGAGAGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100
QY 301 CTGGACAAGATCGAGGAGAGCAGACAAGAGTGCAGCAGAGATCCAGAGGCGGAGGCC 360
Db 101 LeuAspLysIleGluGlnGluGlnAsnLysSerLysLysValAlaGlnGlnAla--Ala 119
QY 361 GCCGACAAAGGCGC-----AGGTAGCCAGAACTACCCCATCTGTGCAGAACCTGCAG 411
Db 120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139
QY 412 GGCACAGATGTGCACCGAGGCATCAGCCCGCGACCCCTGAACGCTGGGTGAAGGTGATC 471
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValVal 159
QY 472 GAGGAGAAAGGCGCTTCAGCCCCGAGGTGATCCCCCATGTTCACCGCCCTGAGCGAGGCGCC 531

```

Db 160 GluGluValaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179  
QY 532 ACCCCCGAGGAGCTGAACACGATGTTGAACACCGTGGCGGCCACAGCGCCCATGACG 591  
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGln 199  
QY 592 ATGCTGAAGGACACCATCAACGAGAGGCGCCGAGTGGGACCGCGTGACCCCGTGAC 651  
Db 200 MetLeuysgluThrIleAsnGluGluAlaGluTrpAspArgValHisProValHis 219  
QY 652 GCGGCGCCCATCGCCCGCCGAGATGCGCGGAGCGCCCGCGGACGACATCGCCGCGCAC 711  
Db 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239  
QY 712 ACCAGCACCTGACGAGAGGATCGCCTGGATGACCAACCCCGCATCCCGTGCGC 771  
Db 240 ThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProProIleProValGly 259  
QY 772 GACATCTACAGCGGTGATCATCTGGCGCTGAACACAGATCGTGGGATGTACAGCCCG 831  
Db 260 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279  
QY 832 GTGAGCATCTTGACATCAAGCAGGCGCCCAAGAGAGCGCTTCGCGACTACGTGACCGC 891  
Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299  
QY 892 TTCTTCAAGACCCCTGCGCGCCGAGCAGACACCCAGAGGTGAAGACTGTGATGACCGAC 951  
Db 300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGlu 319  
QY 952 ACCCTGCTGTGACAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCGCC 1011  
Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339  
QY 1012 GCGCGCAGCTGAGAGGATGATGACCGCTGCGAGGCGGTGGCGGCCCGCCAGCACAAG 1071  
Db 340 AlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyLysProGlyHisLys 359  
QY 1072 GCGCGCGTGTGCGCAGGCGATGAGCGAG--GCCAACACC--AGCGTGATGATGACG 1125  
Db 360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln 379  
QY 1126 AAGAGCACTTCAAGGGCCCCCGCGCATCGTCAAGTCTTCACTGCGGCAAGAGGGC 1185  
Db 380 ArgGlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnLysGlyLysGlu 399  
QY 1186 CACATCGCCCGCAACTGCGCGCGCCCGCAAGAGGGCTGTGAAGTGGCGCAAGAG 1245  
Db 400 HisThrAlaArgAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGlu 419  
QY 1246 GCGCACCATGAAGACTGACCGGAGCGCGCAACTTCTGGGCAAGATC--TGG 1302  
Db 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleCysLeu 439  
QY 1303 CCCAGCCACAAGGCGCGCCCGCAACTTCTGAGAGCGCGCCGAGCCACCGCGCC 1362  
Db 440 ProThrArgGluGln--GlyIlePhePhe-----ArgAlaAspGlnSerGlnGln 456  
QY 1363 CCCGCGGAGAGCTTCGCTTCGAGAGACC--ACCGCGCGCGCAAGAGCAGAGAGC 1416  
Db 457 ProHisHisPhePheArgAlaAspGlnSerGlnGlnProHisGlnLysArgAlaSer 475

RESULT 15  
US-08-392-794A-2  
Sequence 2, Application US/08392794A  
Patent No. 6025141  
GENERAL INFORMATION:  
APPLICANT: HU, Yu-Wen  
TITLE OF INVENTION: IMMUNOFLUORESCENCE ASSAY FOR THE DETECTION OF  
TITLE OF INVENTION: ANTIBODIES USING RECOMBINANT ANTIGENS IN INSOLUBLE FORM  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,794A  
FILING DATE: 09-DEC-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/164,789  
FILING DATE: 10-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: NEIMARK, SHERIDAN  
REGISTRATION NUMBER: 20,520  
REFERENCE/DOCKET NUMBER: HU=4A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-392-794A-2

Alignment Scores:  
Pred. No.: 4.38e-116 Length: 437  
Score: 2023.00 Matches: 376  
Percent Similarity: 94.06% Conservative: 36  
Best Local Similarity: 85.84% Mismatches: 20  
Query Match: 72.98% Indels: 6  
DB: 3 Gaps: 3

US-09-475-704A-3 (1-1479) x US-08-392-794A-2 (1-437)

QY 1 ATGGCGCGCGCGCCAGCATCTCGCGCGGCAAGCTGACGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGCGCAGCGCAG 120  
Db 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
QY 121 CTGAGAGTTCGCGCTGAAACCCCGCGCTGTGAGACCAAGAGCTGAGAGCTGTTCAAC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
QY 181 ATCCGCAAGCTGCACCCCGCGCTGACAGCGGCGAGCGAGAGCTGAAGAGCTGTTCAAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCAGCTGTACTGCTGTGACGAGAGAGATCGAGCTCGCGACCAAGAGAGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100  
QY 301 CTGACACAAGATCGAGAGGAGCAGAACAGTGCACAGAGAGATCCAGAGCGCGAGCC 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysAlaGlnGlnAla--Ala 119  
QY 361 GCGGACAAGGCGC-----AAGTGAGCCAGAACTACCCCATCGTGACAGAACCTGACG 411  
Db 120 AlaAspThrGlyHisSerAsnGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139  
QY 412 GCGCAGATGTGACACGAGCCATCAGCCCGCGCACCTGAAGCGCTGGGTGAAGTGATC 471

```
Db      140  ||||| 140 GlyIleMetValHisGlnIleIleSerProArgThrLeuAsnAlaTrpValIysValVal 159
QY      472  ||||| 472 GAGGAGAAGGCTTACGCGCGAGGTGATGCCATGTTACCGCCCTGAGCGAGGCGCC 531
Db      160  ||||| 160 GluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179
QY      532  ||||| 532 ACCCCCAAGACCTGAACACGATGTTGAACCGTGCGCGCCACCGAGCGCCCATGAC 591
Db      180  ||||| 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaIleMetGln 199
QY      592  ||||| 592 ATGCTGAAGGACACCATCAACGAGAGCGCGCGAGTGGACCGCGTGCACCCCGTGAC 651
Db      200  ||||| 200 MetLeuLysGluThrIleAsnGluGluAlaIleGluTrpAspArgValHisProValHis 219
QY      652  ||||| 652 GCCGGCCCATCGCCCGCGCGAGATGCGCGAGCCCGCGCGACGACATCGCCGCGAC 711
Db      220  ||||| 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239
QY      712  ||||| 712 ACCAGCACCTTGACGAGACGATCGCTGATGACCAACACCCCGCATCCCGCTGGGC 771
Db      240  ||||| 240 ThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProProIleProValGly 259
QY      772  ||||| 772 GACATCTCAAGCGGTGATCATCTGCGCTGAACAAGATCGTGCGATGTACAGCCCC 831
Db      260  ||||| 260 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279
QY      832  ||||| 832 GTGAGCATCTTGACATCAAGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGACCGC 891
Db      280  ||||| 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299
QY      892  ||||| 892 TTCTTCAAGACCTGCGCGCGAGCAGACGACCCAGAGGTGAAGACTGATGACCGAC 951
Db      300  ||||| 300 PheTyrLysThrLeuArgAlaGlnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 319
QY      952  ||||| 952 ACCCTGCTGCTGACGAAACCCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCCCC 1011
Db      320  ||||| 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339
QY      1012  ||||| 1012 GCGCGCACCTTGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCAAGCAACAAG 1071
Db      340  ||||| 340 AlaAlaThrLeuGluGlnMetMetThrAlaCysGlnGlyValGlyProGlyHisLys 359
QY      1072  ||||| 1072 GCCCGCTGCTGGCGCGAGCGATGAGCCGACCAACACCGAC-----GTGATGATGCAG 1125
Db      360  ||||| 360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGln 379
QY      1126  ||||| 1126 AAGAGCAACTTCAAGGGCGCGCGCATCGTCAAGTGTCAACTGCGGCAAGAGGGC 1185
Db      380  ||||| 380 ArgGlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCysGlyLysGluGly 399
QY      1186  ||||| 1186 CACATCGCGCGCAACTGCGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 1245
Db      400  ||||| 400 HisThrAlaArgAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGlu 419
QY      1246  ||||| 1246 GGGCACCAAGATGAAGGACTGCACCGAGCGCCAGGCGCAACTTCTGGGCAAGATC 1299
Db      420  ||||| 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIle 437
```

Search completed: March 11, 2005, 15:45:52  
Job time : 73.8258 secs

**SHIFT  
EAGE BLANK (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 14:56:14 ; Search time 202.694 Seconds  
(without alignments)  
5644.151 Million cell updates/sec

Title: US-09-475-704A-3  
Perfect score: 2772  
Sequence: 1 atggcgccgcgcgcagcat.....acgacccctgagccagtaa 1479

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO\_spool\_p/US09475704/runat\_10032005\_140220\_14808/app\_query.fasta\_1.3342  
-DB=A\_Geneseq\_16Dec04 -QFMT=faстан -SUPFX=n2p.rag -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09475704 @CGN\_1\_1\_472 @runat\_10032005\_140220\_14808 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_16Dec04:\*

1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2606	94.0	492	3	AA96943 HIV synth
2	2448	88.3	492	6	AAE37600 HIV-1 sub
3	2441	88.1	492	8	ADP02848 wild type
4	2441	88.1	513	8	ADP02847 wild type
5	2436.5	87.9	514	8	ADN36396 Human pro
6	2435	87.8	492	5	AAE18303 Human imm
7	2435	87.8	635	5	AA48948 HIV-1 sub
8	2423	87.4	492	7	ADC72880 HIV-1 gag
9	2419.5	87.3	1457	8	ADN36406 Human pro
10	2416	87.2	492	7	ADC72879 HIV-1 gag

11	2416	87.2	494	3	AAB69275	Aab69275 HIV-1 non
12	2415.5	87.1	491	7	ADC72910	Adc72910 HIV-1 gag
13	2407.5	86.9	491	7	ADC72874	Adc72874 HIV-1 gag
14	2405.5	86.8	491	3	AAB69278	Aab69278 HIV-1 non
15	2396.5	86.5	491	7	ADC72913	Adc72913 HIV-1 gag
16	2383	86.0	498	3	AAB69276	Aab69276 HIV-1 non
17	2382	85.9	496	7	ADC72878	Adc72878 HIV-1 gag
18	2381.5	85.9	487	7	ADC72876	Adc72876 HIV-1 gag
19	2381.5	85.9	497	7	ADC72877	Adc72877 HIV-1 gag
20	2344	84.6	508	7	ADC72875	Adc72875 HIV-1 gag
21	2343	84.5	502	3	AA96944	Aay96944 HIV synth
22	2341.5	84.5	3025	4	AAB69271	Aab69271 HIV-1 sub
23	2329	84.0	494	3	ADC72896	Adc72896 HIV-1 non
24	2328.5	84.0	494	7	ADC72896	Adc72896 HIV-1 gag
25	2301	83.0	500	1	AA980884	Aap80884 Sequence
26	2297	82.9	500	7	ADC72788	Adc72788 HIV-1 gag
27	2294.5	82.8	497	2	AA980840	Aar08040 Sequence
28	2277	82.1	498	8	ADO52538	Ado52538 HIV-1 rec
29	2269	81.9	500	6	ABR55488	Abri55488 Amino aci
30	2267	81.8	500	3	AAB69270	Aab69270 HIV-1 non
31	2266.5	81.8	505	1	AA980885	Aap80885 Sequence
32	2266.5	81.8	505	2	AAW72992	Aaw72992 HIV isolat
33	2266	81.7	500	8	ADO52524	Ado52524 HIV-1 rec
34	2263	81.6	500	8	ADO52550	Ado52550 HIV-1 rec
35	2262	81.6	498	8	ADO52543	Ado52543 HIV-1 rec
36	2260	81.5	508	1	AA960419	Aap60419 Sequence
37	2259.5	81.5	492	3	AAB69268	Aab69268 HIV-1 non
38	2258	81.5	500	7	ADC72891	Adc72891 HIV-1 gag
39	2258	81.5	508	5	AAO19386	Aao19386 Lymphaden
40	2256	81.4	500	2	AAW30562	Aaw30562 HIV-1 GAG
41	2256	81.4	500	2	AAW30569	Aaw30569 HIV-1 gag
42	2256	81.4	500	6	ABR44737	Abri44737 Plasmid p
43	2256	81.4	500	6	AAE32734	Aae32734 Human imm
44	2256	81.4	500	7	ADB81007	Adb81007 RING-SH C
45	2256	81.4	500	7	ADF60702	Adf60702 Amino aci

#### ALIGNMENTS

RESULT 1	AA96943	standard; protein; 492 AA.
ID	AA96943	
AC	AA96943;	
DT	31-OCT-2000	(first entry)
XX		
DE	HIV synthetic Gag polypeptide.	
XX		
KW	Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;	
KW	DNA immunization; packaging cell line; antigen presentation.	
XX		
OS	Human immunodeficiency virus; type C strain AF110965.	
OS	Synthetic.	
XX		
PN	WO200039304-A2.	
XX		
PD	06-JUL-2000.	
XX		
PF	30-DEC-1999;	99WO-US031273.
XX		
PR	31-DEC-1998;	98US-0114495P.
PR	01-SEP-1999;	99US-0152195P.
XX		
PA	(CHIR ) CHIRON CORP.	
XX		
PI	Barnett S, Zur Megede J;	
XX		
DR	WPI; 2000-452401/39.	
DR	N-PSDB; AAA51609.	
XX		
PT	Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env polypeptide and the polypeptide useful for immunizing a mammal especially	

PT human against HIV.

XX XX Disclosure; Page 101-103; 113pp; English.  
XX XX Expression cassettes comprising a polynucleotide encoding antigenic type  
CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
CC in DNA immunization, generation of packaging cell lines and production of  
CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
CC cassettes exhibit increased potency for induction of cytotoxic T-  
CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-  
CC assemble into non-infectious virus-like particles which are used as a  
CC matrix for the proper presentation of an antigen entrapped or associated  
CC to the immune system of the host

XX SQ Sequence 492 AA;

Alignment Scores:

Pred. No.: 4.04e-147 Length: 492  
Score: 2606.00 Matches: 492.  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 94.01% Indels: 0  
DB: 3 Gaps: 0

US-09-475-704A-3 (1-1479) x AAY96943 (1-492)

QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspAlaTrpIuArgIleArg 20  
QY 61 CTGCGCGCGCGCGCGCAAGAGTGTATGATGAGCACTGTGTGGCGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysCysTyrMetMetLysHisLeuValTrpAlaSerArgIu 40  
QY 121 CTGAGAGAGTTCGCCCTGAACCCCGGCTGTGAGACCGAGCGGCTGCAAGAGATC 180  
Db 41 LeuGluLysPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATCGCGCCAGCTGCACCCCGCGCTGAGACCGGCGAGGAGGAGCTGAAGCCCTGTTCAAC 240  
Db 61 IleArgGlnLeuHisProAlaLeuGlnThrGlySerGluGluLeuLysSerLeuPheAsn 80  
QY 241 ACCGTGGCCACCCCTGTACTGCGTGCAGAGAGATGAGGTCCGCGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAGGAGGAGGAGCAACAAGTGCAGAGAGATCCAGCGCGAGGCC 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysIleGlnGlnAlaGluAla 120  
QY 361 GCCGACAAGGGGCAAGGTGAGCCAGAACTACCCCATGTGCAGAACTGCGAGGGCGAGATG 420  
Db 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 140  
QY 421 GTGCAACGAGGCGCATCAGCCCGCGCACCCCTGAACGCGCTGGTGAAGGTGATCGAGAGAG 480  
Db 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 160  
QY 481 GCGTTCAAGCCCGGAGGTATCCCATGTTCAACGCGCTGAGCGAGGGCGCGCACCCCGCAG 540  
Db 161 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 180  
QY 541 GACCTGAACACGATGTGAACACCGTGGCGGCGGCGCACGCGCGCATGAGATGCTGAAG 600  
Db 181 AspleuAsnThrMetLeuAsnThrValGlyHisGlnAlaIleMetGlnMetLeuLys 200  
QY 601 GACACCATCAAGAGGAGGCGCGGAGTGGAGCGCGGTGCACCCCGTGCAGCGCGGCGCCC 660  
Db 201 AspThrIleAsnGluGluAlaIleArgIuTrpAspArgValHisProValHisAlaGlyPro 220  
QY 661 ATGCGCGCGCGCGAGATGCGCGAGCGCGCGCGCGAGCAGCATCGCGCGGCAACGAGCACCC 720  
Db 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240

QY 721 CTGCAAGAGCAGATCGCCTGTGATGACCAGCAACCCCGCATCCCGGTGGCGCATCTAC 780  
Db 241 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyr 260  
QY 781 AAGCGGTGATCATCTCTGGGCTGAACAAGATGTGGGATGTACAGCCCGTGAAGCATC 840  
Db 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 280  
QY 841 CTGGAACATCAAGAGGCGCGCGCAAGAGCCCTTCGCGACTACGTGAGCCGCTTCTCAAG 900  
Db 281 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 300  
QY 901 ACCCTGCGCGCGCGAGCAGACGACCCAGAGGTGAAGAAGTGAATGACCGACACCTGCTG 960  
Db 301 ThrLeuArgAlaGluGlnSerThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 320  
QY 961 GTGCAAGACCGCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCGCGCGCGCAGC 1020  
Db 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340  
QY 1021 CTGGAAGAGATGATGACCGCGCTGCGCAGGCGCGTGGCGCGCGCGCGCGCGCTG 1080  
Db 341 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 360  
QY 1081 CTGGCGCGAGCGATGAGCGCGCGCAACACCGCGTGTATGATGCAAGAGCAACTCAAG 1140  
Db 361 LeuAlaGluAlaMetSerGlnAlaAsnThrSerValMetMetGlnLysSerAsnPheLys 380  
QY 1141 GGCGCGCGCGCGCATGCTCAAGTGTCTCAACTGCGCGCAAGAGGCGCGCATGCGCGCAAC 1200  
Db 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 400  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGCTGTGGAAGTGGCGCAAGAGGCGCGCACGATGAAG 1260  
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420  
QY 1261 GACTGCACCGAGCGCGCGCAAGGCGCAACTCTCGGCGCAAGATCGGCGCGCACAGGCGCGC 1320  
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyIleTrpProSerHisLysGlyArg 440  
QY 1321 CCGCGCAACTCTCTGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1380  
Db 441 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 460  
QY 1381 TTCGAGAGAGCAACCCCGCGCGCAAGAGCGAGAGCAAGAGCGCGAGACCTTGACGAGC 1440  
Db 461 PheGluGluThrThrProGlyGlnLysGlnGluSerLysAspArgGluThrLeuThrSer 480  
QY 1441 CTGAAGAGCGCTGTTCGCAAGCAGACCCCGCTGAGCCAG 1476  
Db 481 LeuLysSerLeuPheGlyAsnAspProLeuSerGln 492

RESULT 2

AAE37600 standard; protein; 492 AA.

AAE37600;

23-OCT-2003 (revised)  
27-AUG-2003 (first entry)

HIV-1 subtype C isolate Du422 Gag protein.

Regulatory gene; accessory gene; HIV; human immunodeficiency virus;

vacine; infection; gene therapy; Gag.

Human immunodeficiency virus 1.

WO2003037919-A2.

08-MAY-2003.

PF 31-OCT-2002; 2002WO-IB004550.  
XX  
PR 31-OCT-2001; 2001ZA-00008978.  
XX  
PA (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.  
PA (UYCA-) UNIV CAPE TOWN.  
XX  
PI Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;  
XX  
DR WPI: 2003-430497/40.  
DR N-PSDB; AAD29257.  
XX  
PT New molecules comprising HIV-1 subtype isolate regulatory/accessory  
PT genes, useful for manufacturing a vaccine for treating or preventing HIV  
PT infection.  
XX  
PS Disclosure; Page 91-92; 97pp; English.  
XX  
CC The invention relates to molecules comprising HIV-1 subtype isolate  
CC regulatory/accessory genes (tat, nef and rev genes) and modifications and  
CC derivatives thereof. The invention also provides proteins encoded by such  
CC genes. Sequences of the invention are useful for manufacturing vaccines  
CC for treating or preventing human immunodeficiency virus (HIV) infections.  
CC They are also useful in gene therapy. The present sequence is HIV-1  
CC subtype C isolate Du422 Gag protein. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 492 AA;  
  
Alignment Scores:  
Pred. No.: 1.05e-137 Length: 492  
Score: 2448.00 Matches: 461  
Percent Similarity: 97.36% Conservative: 19  
Best Local Similarity: 93.51% Mismatches: 11  
Query Match: 88.31% Indels: 2  
DB: 6 Gaps: 2  
  
US-09-475-704a-3 (1-1479) x AAE37600 (1-492)  
QY 1 ATGGGGCGCCGCGCCAGATCCTGCGCGCGGCAAGCTGGAAGCCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGlyLysIleArg 20  
QY 61 CTGCGCGCGCGCGGCAAGAGTGTCTATGATGAAGACCTGTGTGGCCAGCCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisIleTyrMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAAGAAGTTCGCTGAACCCCGGCTGTGAGACGAGCGAGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATCCGCGAGCTGCACCCCGCCCTGCAGACCGGCGAGCGAGAGCTGAAGCCTGTTCAC 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGAGAGATCGAGTCCGCGACCAAGAGAGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGlnAla 100  
QY 301 CTGCAAGAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCC 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCGCAAGAGGCAAGGTGAGCCAGAACTAATCCCATCTGCAGAACTGCAGAGCGCCAGATG 420  
Db 121 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMet 139  
QY 421 GTGACCAAGGCGCATCAGCCCGCAGACCTGAAGCGCTGGTGAAGGTGATCGAGAGAG 480  
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159  
QY 481 GCCTTCAGCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCGCAG 540  
|||||

Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179  
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGGCGCCACAGGCGCCATGAGATGCTGAG 600  
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGCACCCCGTGCAGCGCGGCCCC 660  
Db 200 AspThrIleAsnGluGlnAlaAlaGluTrpAspArgValHisProValHisAlaGlyPro 219  
QY 661 ATGCGCGCGCGCGAGATGCGGAGCGCGCGCGCGAGAGATGCGCGCGCACACAGACCC 720  
Db 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGCAAGAGCAGATCGCCTGTGATGACCAACCCCATCCCGCTGGCGGACATGTAC 780  
Db 240 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGGGCTGGAACAAGATCGTGGATGTACAGCCCGTGAGCATC 840  
Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279  
QY 841 CTGCAATCAAGCAGGCGCGCGCAAGAGCCCTTCCGCGACTACGTGGAACGCTTCTTCAAG 900  
Db 280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGCGAGAGAGACCAAGAGGTGAAGAACTGATGACCGACACCTGCTG 960  
Db 300 ThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCGGCGCGCGCGAGC 1020  
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339  
QY 1021 CTGGAAGAGATGATGACCGCTGCGAGGCGCGTGGGCGCGCGCGCGCGCGCTG 1080  
Db 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgVal 359  
QY 1081 CTGCGCGAGCGCATGAGCCAGCGCAACAC--AGCGTATGATGAGAGAGCAACTTC 1137  
Db 360 LeuAlaGluAlaMetSerGlnAlaAsnSerGlyAsnIleMetMetGlnArgSerAsnPhe 379  
QY 1138 AAGGCGCGCGCGCGCATGTCAGAGCTTCAACTGCGCGCAAGAGGCGCCACATCGCCGC 1197  
Db 380 LysGlyProArgArgGlyValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
QY 1198 AACTGCGCGCGCGCGCGCGCAAGAGGCTGCTGAAGTGGCGCAAGAGGCGCCACAGATG 1257  
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleGlnMet 419  
QY 1258 AAGCATGCAACCGCGCGCGCAACTTCTGCGCAAGATCTGCGCGCGCGCGCGCGAGCTTC 1317  
Db 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439  
QY 1318 CGCGCGCGCGCACTTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGAGCTTC 1377  
Db 440 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTCGAGAGAGACACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
Db 460 ArgPheGluGluThrThrProAlaProLysGlnGluProIleGluArgGluProLeuThr 479  
QY 1438 AGCTTGAAGAGCTGTCGCGCAAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476  
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492  
  
RESULT 3  
ADP02848 standard; protein; 492 AA.  
ID ADP02848  
XX ADP02848;  
AC  
XX  
DT 18-NOV-2004 (first entry)







```
QY 1141 GGGCCCCGGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCCGCAC 1200
    ||| ::|||
Db 403 GlySerLysArgIleValLysCysPheAsnCySGlyLysGluGlyHisIleAlaArgAsn 422
QY 1201 TGCCGGCCCCCGGCAAGAGGGCTGCTGGAAGTGGCGGCAAGAGGGCCACGATGAAG 1260
    |||
Db 423 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 442
QY 1261 GACTGCACCGAGCGCGGCGCAACTTCTGCGGCAAGATCTGGCCCAAGCCCAAGGGCCGC 1320
    |||
Db 443 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 462
QY 1321 CCGGCGCAACTTCTGCGAGAGCGCGCGGAGCCCGCGCGCGCGCGCGGAGAGCTTCCGC 1380
    |||
Db 463 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGlnSerPheArg 482
QY 1381 TTGAGAGAGACCAACCCCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
    |||
Db 483 PheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuThrSer 502
QY 1441 CTGAAGAGCTGTTCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
    |||
Db 503 LeuLysSerLeuPheGlySerAspProLeuSerGln 514

RESULT 6
AAE18303
ID AAE18303 standard; protein; 492 AA.
AC AAE18303;
XX 29-AUG-2003 (revised)
DT 07-MAY-2002 (first entry)
XX
DE Human immunodeficiency virus type 1 (HIV-1) Clade C gag protein.
XX
KW Human immunodeficiency virus; vaccine; HIV infection; immune response;
KW alpha-virus replicon; therapy; gag gene.
XX
OS Human immunodeficiency virus 1.
XX
PN WO200203917-A2.
XX
PD 17-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US021701.
XX
PR 07-JUL-2000; 2000US-0216995P.
XX
PA (ALPH-) ALPHAVAX INC.
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Olmedo R, Keith P, Dryga S, Caley I, Maughan M, Johnston R;
PI Davis N, Swanson R;
XX
DR WPI; 2002-171664/22.
DR N-PSDB; AAD29133.
XX
PT Composition useful for treating or preventing HIV infections, comprises
PT two or more isolated nucleic acids encoding env, gag or pol gene product
PT of HIV or immunogenic fragment of the gene products.
XX
PS Example 5; Page 165-166; 201p; English.
XX
CC The invention relates to a composition comprising isolated nucleic acids
CC encoding env, gag or pol gene product of human immunodeficiency virus or
CC immunogenic fragment of the gene products. The gag gene product is
CC modified to inhibit formation of virus-like particles containing gag gene
CC product and their release from cells, and the pol gene product is
CC modified to inhibit reverse transcriptase activity. The invention also
CC relates to a method for producing an alpha-virus replicon particle used
CC in vaccines. The composition is useful for inducing an immune response to
CC human immunodeficiency virus (HIV) or for treating or preventing HIV
CC infection in a subject. The alpha-virus replicon particle is useful in a
```

```
CC vaccine. The composition is useful for administering a protein or peptide
CC to a subject. A composition comprising heparin affinity-purified alpha-
CC virus replicon particle is useful as a clinical trial material and as a
CC commercial product. The present sequence is HIV-1 Clade C gag protein.
CC (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 492 AA;

Alignment Scores:
Pred. No.: 6.28e-137 Length: 492
Score: 2435.00 Matches: 458
Percent Similarity: 96.96% Conservative: 20
Best Local Similarity: 92.90% Mismatches: 13
Query Match: 87.84% Indels: 2
DB: 5 Gaps: 2

US-09-475-704A-3 (1-1479) x AAE18303 (1-492)
QY 1 ATGGGGCGCGCGCGCATCTGCGCGGCGGCAAGCTGGAGCGCTGGAGCGCATCGCC 60
    |||
Db 1 MetaAlaAlaArgAlaSerIleLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20
QY 61 CTGCGCGCGCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGCCAGCGCGAG 120
    |||
Db 21 LeuArgProGlyGlyLysLysHisIleTyMetLeuLysHisIleValTrpAlaSerArgGlu 40
QY 121 CTGGAAGATTGCGCCCTGAACCCCGGCTGCTGAGAGACGAGGAGGCTGCAAGCAGATC 180
    |||
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
QY 181 ATCCGCGCATGTGACACCCCGCGCTGACAGCGGAGAGAGAGCTGAAGAGCTGTTCAAC 240
    |||
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyAsn 80
QY 241 ACCGTGCGCACCCCTGTACTGCGTGCAGAGAGAGAGAGAGTCCGCGACCAAGAGAGCC 300
    |||
Db 81 ThrValAlaThrLeuTyrcysValHisGluLysIleGluValArgAspThrLysGluAla 100
QY 301 CTGCAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCC 360
    |||
Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120
QY 361 GCCGACAAAGGCAAGGTGAGCCAGAGACTACCCCATCTGTGAGAACTGACAGGCGCCAGTG 420
    |||
Db 121 AlaAsp--GlyLysValSerGlnAsnTyrcProIleValGlnAsnLeuGlnGlyGlnMet 139
QY 421 GTGACCAAGGCATCAGACCCCGCGACCTGAAAGCCTGGGTGAAGGTGATCGAGAGAGAG 480
    |||
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159
QY 481 GCCTTCAGCCCGGAGGTGATCCCACTGTTCAACCGCCTGAGCGAGGCGCCACCCCGCAG 540
    |||
Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179
QY 541 GACCTGAACAGATGTGAACACCGGTGGGGCGGCAAGCGCGGCATGCGAGATGCTGAAG 600
    |||
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199
QY 601 GACACCATCAACGAGAGAGCGCGCGAGGTGGAGCCGCGTGACACCCCGTGACAGCGCGCCC 660
    |||
Db 200 AspThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 219
QY 661 ATGCCCCCGGCGGAGATGCGGCGGAGCGCGCGGCGGAGCATCGCGGCAACCAAGAGAGCC 720
    |||
Db 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239
QY 721 CTGCAAGAGCAGATCGCCTGTGATGACCAAGACCCCGCATCCCGTGGGGCGAGCATCTAC 780
    |||
Db 240 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyrc 259
QY 781 AAGCGGTGATCATCTGCGGCTGAACAAGATCGTCGGAGATGTACAGCCCGGTGAGCATC 840
    |||
Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrcSerProValSerIle 279
```

```
QY      841 CTGGACATCAAGCAGGGCCCAAGAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAG 900
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299
QY      901 ACCCTGGCGCCGAGCAGACGACCCAGAGGTTGAAGAACTGATGACCAACCTGCTG 960
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      300 ThrLeuArgAlaGlnAlaThrGlnGlnValLysAsnTrpMetThrAspThrLeuLeu 319
QY      961 GTGCAGAACGCCAACCCGCACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCCAGC 1020
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339
QY      1021 CTGAGAGAGATGATGACCGCTGCAAGGCGTGGCGGCCCAAGGAGGCGCATCGCCGC 1080
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      340 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyProGlyHisLysAlaArgVal 359
QY      1081 CTGGCCGAGCGCATGAGCGCAAGGCCAACAC--AGCGTATGATGCAAGAGCAACTTC 1137
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      360 LeuAlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPhe 379
QY      1138 AAGGCGCCCGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGCGCACATCGCCGC 1197
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaArg 399
QY      1198 AACTGCGCGCGCCCGCCCAAGAGGGCTGTGGAAGTGGCGCAAGAGGCGCCACCAAGTG 1257
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisGlnMet 419
QY      1258 AAGGATGCAACGAGCGCGCAAGCCAACTCTGGAAGATGTGGCCAGCCACCAAGGCG 1317
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      420 LysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439
QY      1318 CGCCCGCGCAACTCTCTGCAAGAGCCGCGCGCAAGCCCGCCCGCGCGCGAGAGCTTC 1377
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      440 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 459
QY      1378 CGCTTCGAGAGAGACCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      460 ArgPheGlnGlnThrThrProAlaProLysGlnGlnProIleGlnArgGluProLeuThr 479
QY      1438 AGCTGAAGAGCTGTTCGGAACGAGACCCCTGAGCCAG 1476
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 7
AAM48948
ID      AAM48948 standard; protein; 635 AA.
XX
AC      AAM48948;
XX
XX      29-AUG-2003 (revised)
DT      19-APR-2002 (first entry)
XX
DE      HIV-1 subtype C isolate Du422 gag protein.
XX
XX      HIV-1 subtype C; vaccine; HIV infection; AIDS; gag; antiviral.
XX
OS      Human immunodeficiency virus; type I.
XX
FH      Key
FT      Misc-difference 523 Location/Qualifiers
FT      Misc-difference 526 /note= "encoded by TAA"
FT      Misc-difference 526 /note= "encoded by TAG"
FT      Misc-difference 557 /note= "encoded by TAA"
FT      Misc-difference 612 /note= "encoded by TGA"
FT
XX
XX      WO200204494-A2.
XX
PD      17-JAN-2002.
```

```
XX
PF      09-JUL-2001; 2001WO-IB001208.
XX
XX      07-JUL-2000; 2000US-0216995P.
PR      10-JUL-2000; 2000ZA-00003437.
PR      15-SEP-2000; 2000ZA-00004924.
XX
PA      (MEDI-) MEDICAL RES COUNCIL.
PA      (UYCA-) UNIV CAPE TOWN.
PA      (UYNC-) UNIV NORTH CAROLINA.
XX
PI      Williamson C, Swanstrom RI, Morris L, Karim SA, Johnston RE;
XX
DR      WPI; 2002-171700/22.
DR      N-PSDB; AAL41591.
XX
PT      Selecting HIV-1 subtype C isolates, which are useful in developing
PT      vaccines against HIV infection, comprises isolating viruses with high
PT      sequence identity to a consensus sequence whose phenotype is associated
PT      with the HIV subtype.
XX
XX      Claim 17; Page 64-66; 69pp; English.
XX
CC      The present invention relates to a process for the selection of human
CC      immunodeficiency virus (HIV) subtype isolates for use in the development
CC      of a prophylactic and/or therapeutic pharmaceutical composition. The
CC      process involves selecting isolated virus or viruses with a high sequence
CC      identity to a consensus sequence and a phenotype which is associated with
CC      transmission for the particular HIV subtype. The composition can be used
CC      in the production of vaccines against HIV. The present sequence is the
CC      HIV-1 subtype C isolate Du422 gag protein. (Updated on 29-AUG-2003 to
CC      standardise OS field)
XX
SQ      Sequence 635 AA;
XX
XX
Alignment Scores:
Pred. No.: 6,48e-137 Length: 635
Score: 2435.00 Matches: 458
Percent Similarity: 96.96% Conservative: 20
Best Local Similarity: 92.90% Mismatches: 13
Query Match: 87.84% Indels: 2
DB: 5 Gaps: 2

US-09-475-704A-3 (1-1479) x AAM48948 (1-635)
QY      1 ATGGGCGCGCGCCGAGCATCTGCGCGCGCAAGCTGAGCGCCTGGAGCGCATCCGC 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      31 MetAlaAlaArgAlaSerIleLeuArgGlyGlnLysLeuAspLysTrpGlnLysIleArg 50
QY      61 CTGCGCGCGCGCGCGCAAGAGTGTCAATGATGAAGCAACCTGTTGGGCCCGCGAG 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      51 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisIleValTrpAlaSerArgGlu 70
QY      121 CTGAGAGAGTTCGCCCTGAACCCCGCGCTGCGAGAGAGAGAGAGAGAGAGAGATC 180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      71 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnGlyCysLysGlnIle 90
QY      181 ATCCGCCAGCTGACCCCGCGCTGACAGACCGGACGAGAGACTGAAGAGCTGTCAAC 240
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      91 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuLysSerLeuTyrAsn 110
QY      241 ACCGTGGCCACCTGTACTGCTGCACAGAGAGATCGAGTCCGCGACACCAAGAGGCC 300
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      111 ThrValAlaThrLeuTyrCysValHisGlnLysIleGlnValArgAspThrLysGlnAla 130
QY      301 CTGAGACAAGATCGAGAGAGAGAGAGACAAGTCCAGAGATCCAGAGCGCGAGGCC 360
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      131 LeuAspLysIleGlnGlnGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 150
QY      361 GCCGACAAGGCGAAGGTGAGCGCAGACACTACCCATCGTGCAGAACCTGCAGGGCCAGATG 420
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      151 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 169
```

```
QY 421 GTGCACCAAGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGTGATCGAGAGAAC 480
Db 170 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValIleGluGluIys 189
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGGGCCCCACCCCCAG 540
Db 190 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 209
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCCCATGAGATGCTGAAG 600
Db 210 AspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuIys 229
QY 601 GACACCATCAACGAGAGAGCCCGCGAGTGGGACCGCGTGCAACCCCGTGACCGCGGCC 660
Db 230 AspThrIleAsnGluGlnAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 249
QY 661 ATGCCCCCGCCAGATGCGCGGAGAGCCCGCGGAGCGGACATCGCGGACCCAGCACACC 720
Db 250 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 269
QY 721 CTGCAGAGACGATCGCGCTGGATGACCAACCAACCCCGCATCCCGTGCGGACATCTAC 780
Db 270 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyr 289
QY 781 AAGCGGTGGATCATCTCTGGGCTGAACACAGATCGTGGGATGTACAGCCCGCTGAGCATC 840
Db 290 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 309
QY 841 CTGCACATCAAGCAGGCGCCCAAGGAGCCCTTCGCGCACTACGTGACCGCTTCTTCAAG 900
Db 310 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheIys 329
QY 901 ACCCTGCGCGCGAGCAGAGCACCCAGAGGTGAAGACTGATGACCGACACCCCTGCTG 960
Db 330 ThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 349
QY 961 GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCGCGGCCAGC 1020
Db 350 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 369
QY 1021 CTGAGAGATGATGACCGCTGCGGCGGTGGCGGCCCGCCAGCCACAAAGCCCGCTG 1080
Db 370 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyIleProGlyHisLysAlaArgVal 389
QY 1081 CTGCGCAGGCGGATGAGCCAGGCCAACAC--AGCGTATGATGCAAGAAGCAACTTC 1137
Db 390 LeuAlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPhe 409
QY 1138 AAGGCCCCCGCGCATCGTCAAGTCTTCAACTGCGGCAAGAGGSCCATCGCCGCC 1197
Db 410 LysGlyProArgArgIleValLysCysBheAsnCysGlyLysGluGlnHisIleAlaArg 429
QY 1198 AACTGCGCGCGCCCGCCAGAGAAGGGCTGTGAAGTGGCGCAAGAGGGCCACAGATG 1257
Db 430 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisGlnMet 449
QY 1258 AAGGACTGACCGGCGCCAGCAACTTCTGGGCAAGATCTGGCCAGCCAGCAAGGGC 1317
Db 450 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 469
QY 1318 CGCCCCGCGCAACTTCTGCAAGAGCGCGCCGAGCCACCGCCCGCCCGCGAGAGCTTC 1377
Db 470 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 489
QY 1378 CGCTTCAGAGAGACACCCCGCGCAAGAGCAGAGAGCAAGAGCGGAGACCTGAC 1437
Db 490 ArgPheGluGluThrThrProAlaProLysGlnGluProIleGluArgGluProLeuThr 509
QY 1438 AGCTGAAGAGCTGTTGGCAACGACCCCTGAGCCAG 1476
Db 510 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 522
```

RESULT 8

```
AD C72880
ID AD C72880 standard; protein; 492 AA.
XX
AC AD C72880;
XX
DT 18-DEC-2003 (first entry)
XX
DE HIV-1 gag protein containing an NCP sequence, SEQ ID NO 110.
XX
KW complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;
KW human immunodeficiency virus; HIV.
XX
OS Human immunodeficiency virus 1.
XX
PN WO2003060098-A2.
XX
PD 24-JUL-2003.
XX
PF 10-JAN-2003; 2003WO-US000801.
XX
PR 11-JAN-2002; 2002US-0347369P.
XX
PA (ACHI-) ACHILLION PHARM INC.
XX
PI Beuchter D, Hou X, Marlor CW, Rice WG, Yang W;
XX
DR WPI; 2003-646042/61.
XX
PT Method of determining whether a compound inhibits formation of complex
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site
PT oligonucleotide by comparing amount of complex formed in presence/absence
PT of compound.
XX
PS Claim 2; SEQ ID NO 110; 105pp; English.
XX
CC The invention relates to a novel method of determining whether a compound
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)
CC polypeptide and an HIV-psi-site oligonucleotide. The method involves
CC adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and
CC the novel compound and comparing amount of complex formed, with that
CC formed in the absence of the compound, where a decrease in the amount of
CC complex formed in presence of the compound indicates that the compound
CC inhibits complex formation. The method of the invention involves anti-HIV
CC activity. The compounds identified are useful for treating a subject
CC infected with human immunodeficiency virus (HIV) by administering the
CC compound to the subject. This sequence represents an HIV-1 gag protein
CC which contains an NCP sequence of the invention.
XX
SQ Sequence 492 AA;
XX
Alignment Scores:
Pred. No.: 3,26e-136 Length: 492
Score: 2423.00 Matches: 451
Percent Similarity: 96.96% Conservative: 27
Best Local Similarity: 91.48% Mismatches: 13
Query Match: 87.41% Indels: 2
DB: 7 Gaps: 2
US-09-475-704A-3 (1-1479) x AD C72880 (1-492)
QY 1 ATGGGCGCGCGCGCGCATCTGCGCGGCGGCAAGCTGACGCGCTGGAGCGCATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGluLysLeuAspTrpTrpGluLysIleArg 20
QY 61 CTGGCGCGCGCGCGCAAGAAGTGTACATGATGAACACCTGTGTGGCCAGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysCysTyrMetLeuLysHisIleValTrpAlaSerArgGlu 40
QY 121 CTGAGAGAAGTTGCGCCTGAACCCCGCGCTGTGGAAGACGAGGCGCTGCAAGAGATC 180
Db 41 LeuGluArgPheSerLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
QY 181 ATCCGCCAGCTGACCCCGCGCTGCAAGACCGGAGCGAGAGAGCTGAAGACCTTTCAC 240
```

```
Db      61 MetLysGlnLeuGlnProAlaLeuGlnThrGlnGluLeuLysSerLeuTyrAsn 80
QY      241 ACCGTGGCCACCCTGTACTGCGTGCAACGAGATCGAGGTCCGCCACCAAGAGGCC 300
Db      81 ThrValAlaThrLeuPheCysValHisGlnLysIleAlaValArgAspThrLysGlnAla 100
QY      301 CTGACAGATCGAGAGAGACAGAACTAGTCCAGACAGATCCAGAGGCCCGAGGCC 360
Db      101 LeuAspLysIleGlnGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaValAla 120
QY      361 GCCGACAGGGCAAGGTGAGCCAGAACTACCCCATCGTGACAGAACCTGACGGCCAGATG 420
Db      121 AlaAsp---GlyThrValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 139
QY      421 GTGACACGAGCCATCAGCCCGCCGACCCCTGAACGCGCTGGGTGAAGTGATCGAGAGAG 480
Db      140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 159
QY      481 GCCTTCAGCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGGCCACCCCCAG 540
Db      160 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlnAlaThrProGln 179
QY      541 GACCTGAACACGATGTTGAACACCGTGGCGGCCACGAGCGCCCATGACAGATGCTGAAG 600
Db      180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199
QY      601 GACACCATCAACGAGAGGGCCGCGAGTGGGAGCCGCGTGACCCCGTGACCGCCGCCCC 660
Db      200 AspThrIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHisProValHisAlaGlyPro 219
QY      661 ATGCCCCCGGCGCAGATGCGCGAGCCCGCGGCGGACGACATCGCCGCCACCAACGAGACC 720
Db      220 IleAlaProGlnGlnMetArgGlnProArgGlySerAspIleAlaGlyThrThrSerThr 239
QY      721 CTGACAGAGCAGATCGCCTGATGACGACCAACCCCCCATCCCGCTGGCGGACATCTAC 780
Db      240 LeuGlnGlnGlnIleAlaTrpMetThrAsnAsnProProValProValGlyAspIleTyr 259
QY      781 AAGCGGTGATCATCTCTGGGCTGGAACAAGATCGTGGGATGTACAGCCCGCTGAGCATC 840
Db      260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279
QY      841 CTGACATCAAGCAGAGGGCCCCAAGAGGCCCTTCGCGCACTACGTGACCGCTTCTTCAAG 900
Db      280 LeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPheLys 299
QY      901 ACCCTGCGCGCGCAGAGACGACCCGAGAGGTGAAGACTGATGACCGACACCTGCTG 960
Db      300 ThrLeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319
QY      961 GTGCAAGACGCCAACCCCGACTGACAGACCATCTCTGCGCTGCGGCCCGCGCGCAGC 1020
Db      320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339
QY      1021 CTGAGAGATGATGACCGCCTGCGAGGGCGGTGGCGGCCCGACCAAGAGCCCGCTG 1080
Db      340 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359
QY      1081 CTGCGCGAGGCGATGAGCCAGGCC---AACACACGCGTGATGATGACGAGAGCACTTC 1137
Db      360 LeuAlaGlnAlaMetSerGlnThrAsnAsnAlaAsnIleMetMetGlnArgSerAsnPhe 379
QY      1138 AAGGGCCCCCGCGCATGCTCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATCGCCGC 1197
Db      380 LysGlyProArgArgIleIleLysCysPheAsnCysGlyLysGlnGlnHisLeuAlaArg 399
QY      1198 AACTGCGCGCGCCCGCAAGAGGGCTGTGAAAGTGCGGCAAGAGGGCCACCATG 1257
Db      400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlnHisGlnMet 419
QY      1258 AAGGACTGACCGAGCGCCAGGCCAACTTCTGCGCAAGATCTGGCCCGACGACAGGGC 1317
```

```
Db      420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439
QY      1318 CGCCCGCGCAACTTCTTGACAGACCGCCCGAGCCCAACCGCCCGCGGAGAGCTTC 1377
Db      440 ArgProGlnAsnPheLeuGlnAsnArgProGlnProThrAlaProProAlaGlnSerPhe 459
QY      1378 CGCTTCGAGAGACCAACCCCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
Db      460 ArgPheGlnGlnThrThrProAlaProLysGlnGlnProArgGluArgGluProLeuThr 479
QY      1438 AGCCTGAAGAGCCTGTTGCGCAAGACCCCTGAGCCAG 1476
Db      480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 9
ADN36406
ID   ADN36406 standard; protein; 1457 AA.
XX
AC   ADN36406;
XX
DT   15-JUL-2004 (first entry)
XX
DE   Human protein for anti-HIV vaccine.
XX
KW   anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
OS   Homo sapiens.
XX
PN   WO2004035006-A2.
XX
PD   29-APR-2004.
XX
PF   17-OCT-2003; 2003WO-US033112.
XX
PR   18-OCT-2002; 2002US-0419465P.
XX
PA   (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI   Huang Y, Ho DD, Chen Z;
XX
DR   WPI; 2004-348328/32.
DR   N-PSDB; ADN36405.
XX
PT   Nucleic acid vector comprising at least one HIV sequence operably linked
PT   to a promoter and encoding a protein that does not assemble into viral
PT   particles, useful in immunizing a subject against HIV infection.
XX
PS   Disclosure; SEQ ID NO 20; 166bp; English.
XX
CC   The invention relates to a nucleic acid vector comprising at least one
CC   HIV sequence operably linked to a promoter and encoding a protein that
CC   does not assemble into viral particles. The nucleic acid vector is useful
CC   in immunizing a subject against HIV infection. This sequence corresponds
CC   to a peptide used in the invention.
XX
SQ   Sequence 1457 AA;

Alignment Scores:
Pred. No.:      6e-136      Length:      1457
Score:          2419.50     Matches:      455
Percent Similarity: 96.34%   Conservative: 19
Best Local Similarity: 92.48%   Mismatches:  17
Query Match:      87.28%     Indels:       1
DB:              Gaps:       1

US-09-475-704A-3 (1-1479) x ADN36406 (1-1457)

QY      1 ATGGGCGCGCGCGCCAGCATCTCGCGCGCGGCAAGCTGAGCGCCTGGAGCGCATCCGC 60
Db      24 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGlnLysIleArg 43
QY      61 CTGCGCGCGCGCGGCAAGAGTGTACTATGATGAGACACTGTGTGGGCCAGCGCGAG 120
```

Db 44 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisIleValTTrpAlaSerArgGlu 63  
QY 121 CTGAGAGAGTTCCCTGTAACCCCGCTGTGTGAGAGACCGAGGGCTGCAAGCAGATC 180  
Db 64 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnGlyCysLysGlnIle 83  
QY 181 ATCCGCCAGCTGACCCCGCTGACAGACCGGACGAGAGCTGAAGCTGTTCAC 240  
Db 84 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnIleLeuArgSerLeuPheAsn 103  
QY 241 ACCGTGGCCACCTGTACTGCTGTGACGAGAGAGATCGAGGTCCGCGACACCAAGAGGCC 300  
Db 104 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGlnValArgAspThrLysGlnAla 123  
QY 301 CTGACAGAGATCGAGAGAGAGAGCAACAAGTGCAGAGAGATCCAGACGGCCGAGGCC 360  
Db 124 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 143  
QY 361 GCCGACAAAGGCGAAGGTGAGCCAGAACTACCCCATCGTGAGAACTGACGGCCAGATG 420  
Db 144 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 162  
QY 421 GTGACCAAGGCCATCAGCCCCCGCACCCCTGAAGCGCTGGTGAAGGTGATCGAGAGAG 480  
Db 163 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValIleGlnGlnLys 182  
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCTGAGCGAGGGCGCCACCCCCAG 540  
Db 183 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 202  
QY 541 GACCTGAACAAGATGTTGAACAACCGTGGGGCGCCACAGCGCCGCGCATGAGATGCTGAG 600  
Db 203 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 222  
QY 601 GACACCAATCAAGAGAGAGCGCCGCGAGTGGGACCGCGTGACCCCGTGACCGCGGCC 660  
Db 223 AspThrIleAsnGlnGlnAlaIleGlnAlaGlnTyrAspArgValHisProValHisAlaGlyPro 242  
QY 661 ATGCCCCCGCCAGATGCGCGAGAGCGCCCGCGGACGAGATCGCCGCGACCAACAGCAGC 720  
Db 243 IleAlaProGlnGlnMetArgGlnProArgGlySerAspIleAlaGlyThrThrSerThr 262  
QY 721 CTGACAGAGCAGATCGCTGTGATGACCAACACCCCGCATCCCGTGGCGACATCTAC 780  
Db 263 LeuGlnGlyGlnIleAlaTyrMetThrSerAsnProProValProValGlyLysIleTyr 282  
QY 781 AAGCGGTGATCATCTCGGCTTGAACAAGATCGTGCAGATGTACAGCCCGTGAGATC 840  
Db 283 LysArgTyrPheIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 302  
QY 841 CTGACATCAAGCAGAGCGCCCAAGAGAGCGCTTCCGCACTACGTGACCGCTTCTTCAAG 900  
Db 303 LeuAspIleLysGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePheLys 322  
QY 901 ACCCTGCGCGCCGAGCAGACCAACCAAGAGTGAAGAAGTGAAGTGAAGCAGCAGCCTGCTG 960  
Db 323 ThrLeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTyrMetThrAspThrLeuLeu 342  
QY 961 GTGACAGACGCCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCGCCGCGCAGC 1020  
Db 343 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 362  
QY 1021 CTGAGAGAGATGATGACCGCTGCGAGGGGTGGCGGCCCGACCAAGAGCGCCGCTG 1080  
Db 363 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProAsnHisLysValArgVal 382  
QY 1081 CTGCGCGAGGCGATGAGCGACCGCAACACCAAGCGTGTATGCAAGAGCAACTTCAAG 1140  
Db 383 LeuAlaGlnAlaMetSerGlnAlaAsnGlyThrIleLeuMetGlnArgSerAsnPheLys 402  
QY 1141 GGGCCCGCGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCCGCAAC 1200  
Db 403 GlySerLysArgIleValLysCysPheAsnGlyLysGlnGlyHisIleAlaArgAsn 422

QY 1201 TGCCGCGCCCCCGCAAGAGGCTGTGGAAGTGGCGCAAGAGGSCCAGAGATGAG 1260  
Db 423 CysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGlyHisGlnMetLys 442  
QY 1261 GACTGACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCAGCCCAAGAGGCGCC 1320  
Db 443 AspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTyrProPheHisLysGlyArg 462  
QY 1321 CCCGCAACTTCTGCAAGAGCGCCCGAGGCCACCGCCCCCGCCGAGAGCTTCCGC 1380  
Db 463 ProGlyAsnPheLeuGlnSerArgProGlnProThrAlaProProAlaGlnSerPheGly 482  
QY 1381 TTCGAGAGACCAACCCCGCCCAAGAGCAGAGAGCAAGAGACCGCGAGACCTGACCA 1440  
Db 483 PheGlnGlnThrThrProAlaProLysGlnGlnGlnProLysAspArgGlnProLeuThrSer 502  
QY 1441 CTGAAGAGCTGTTCGCAACGAGACCCCTGAGCCAG 1476  
Db 503 LeuLysSerLeuPheGlySerAspProLeuSerGln 514

RESULT 10  
ADCT2879  
ID ADCT2879 standard; protein; 492 AA.  
XX  
AC ADCT2879;  
DT 18-DEC-2003 (first entry)  
XX  
DE HIV-1 gag protein containing an NCP sequence, SEQ ID NO 109.  
XX  
KW complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;  
KW human immunodeficiency virus; HIV.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO2003060098-A2.  
XX  
PD 24-JUL-2003.  
XX  
PF 10-JAN-2003; 2003WO-US000801.  
XX  
PR 11-JAN-2002; 2002US-0347369P.  
XX  
PA (ACHI-) ACHILLION PHARM INC.  
PI Beuchter D, Hou X, Marlor CW, Rice WG, Yang W;  
XX  
XX WPI; 2003-646042/61.  
DR  
XX  
PT Method of determining whether a compound inhibits formation of complex  
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
PT oligonucleotide by comparing amount of complex formed in presence/absence  
PT of compound.  
XX  
XX  
PS Claim 2; SEQ ID NO 109; 105bp; English.  
XX  
XX The invention relates to a novel method of determining whether a compound  
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)  
CC polypeptide and an HIV-psi-site oligonucleotide. The method involves  
CC adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an NCP sequence of the invention.  
XX  
SQ Sequence 492 AA;  
Alignment Scores:

Pred. No.: 8.52e-136 Length: 492  
Score: 2416.00 Matches: 455  
Percent Similarity: 95.74% Conservative: 17  
Best Local Similarity: 92.29% Mismatches: 19  
Query Match: 87.16% Indels: 2  
DB: 7 Gaps: 2  
US-09-475-704A-3 (1-1479) x ADC72879 (1-492)

QY 1 ATGGGCGCGCGCCAGCATCTCGCGCGCGCAAGCTGGAACGCTGGAGCGCATCCGC 60  
Db 1 MetGylAserAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCAAGAAGTGTACATGATGAAGCACTGTGTGGGCCAGCCCGCAG 120  
Db 21 LeuArgProGlyGlyLysLysLysTyrArgLysLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGAATTGCGCCCTGAACCCCGCGCTGTGAGACCAAGGAGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaLeuAsnSerGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60  
QY 181 ATCCGCGAGCTGCACCCCGCGCTGCAGACCGCGCAGGAGAGCTGAAGACCTGTTCAAC 240  
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn 80  
QY 241 ACCGTGCGCACCTGTACTGCGTGCACAGAGAAGATCGAGTCCGCGACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluValArgAspThrLysGluAla 100  
QY 301 CTGAGACAAGATCGAGAGGAGAGCAAGACAAGTGCAGCAAGATCCAGCAAGGCCAGGCC 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysGlu 120  
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAAGAACCTGCAGGGCCAGATG 420  
Db 121 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 139  
QY 421 GTGACACGAGCCATCAGCCCCCGCACCCCTGAACGCGCTGGGTGAAGTGATCGAGAGAAG 480  
Db 140 ValHisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159  
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTTCAACCGCGCTGAGCGAGGCCGCCACCCCCAG 540  
Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179  
QY 541 GACCTGAACACGATGTTGAACACCGTGGGGCGGCCACAGGCCGCCATGCAATGCTGAAG 600  
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAACGAGAGGGCGCGCGAGTGGAGACCGCGTGACCCCGTGACAGCGCGCCCC 660  
Db 200 AspThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 219  
QY 661 ATCGCCCCCGGCGCAGATGCGGAGCGCGCGCGCAGGAGCATCGCGGCACCAACGACACC 720  
Db 220 ValAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGCAAGAGCAGATCGCTGATGAGCAACACCCCCCATCCCGGTGGCGCAGCATCTAC 780  
Db 240 LeuGlnGluGlnIleThrTrpMetThrSerAsnProProValProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCCCGTAGCATC 840  
Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279  
QY 841 CTGACATCAAGAGAGGGCGCCCAAGAGCCCTTCGCGACTACGTGAGACCGCTTCTCAAG 900  
Db 280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGCAGAGCAGACCAAGAGAGGTGAAGAACTGATGACCGACCACTGTGCTG 960  
Db 300 ValLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319

QY 961 GTGAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCGCTCTCGGCCCGCGCCAGC 1020  
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProAlaAlaSer 339  
QY 1021 CTGAGAGATGATGACCCGCTGCCAGGGCGGTGGCGGCCGCCACAGGCCCGCGTG 1080  
Db 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysValArgVal 359  
QY 1081 CTGGCGGAGCGATGAGCCAGGCCAAC--ACCAAGCTGATGATGCAAGAGCACTTC 1137  
Db 360 LeuAlaGluAlaMetSerGlnAlaAsnThrThrAsnIleMetMetGlnLysSerAsnPhe 379  
QY 1138 AAGGCGCGCGCGCATCGTCAAGTGTCTCACTGCGCAAGAGAGGGCCACATCGCCCGC 1197  
Db 380 LysGlyProArgArgThrValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLys 399  
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGCAGCAAGAGAGGCCACAGATG 1257  
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
QY 1258 AAGACTGCACCGCAGCGCCAGGCAACTTCTGGGCAAGATCGGCCACCAAGAGGC 1317  
Db 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439  
QY 1318 CGCCCCGCAACTTCTGCAAGCGCGCCCGCCAGCCCGCGCGCGCGCGAGAGCTTC 1377  
Db 440 ArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTCGAGAGACCAACCCCGCGCCAGAGCAGAGAGCAAGCAAGCAAGCACTGACC 1437  
Db 460 ArgPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuThr 479  
QY 1438 AGCTGAAGAGCTGTTGCGCAACGACCCCTGAGCCAG 1476  
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492  
RESULT 11  
AAB69275  
ID AAB69275 standard; protein; 494 AA.  
XX AC AAB69275;  
XX DT 12-SEP-2003 (revised)  
XX DT 20-APR-2001 (first entry)  
XX DE HIV-1 non-subtype B clone 96ZM651-8 gag protein.  
XX KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
XX KW vif; vpr; tat; rev; nef; vaccine.  
XX OS Human immunodeficiency virus 1.  
XX PN WO200026416-A1.  
XX PD 11-MAY-2000.  
XX PF 25-OCT-1999; 99WO-US024837.  
XX PR 02-NOV-1998; 98US-00184418.  
XX PA (UABR-) UAB RES FOUND.  
XX PI Hahn BH, Shaw GM, Gao F;  
XX DR WPI; 2000-365651/31.  
XX PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.  
XX PS Claim 41; Fig 14; 131pp; English.  
XX CC The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 494 AA;

Alignment Scores:

Pred. No.:	8.52e-136	Length:	494
Score:	2416.00	Matches:	455
Percent Similarity:	96.77%	Conservative:	24
Best Local Similarity:	91.92%	Mismatches:	12
Query Match:	87.16%	Indels:	4
DB:	3	Gaps:	3

US-09-475-704A-3 (1-1479) x AAB69275 (1-494)

```

QY 1 ATGGGCGCCGCGCCAGCATCTGCGCGCGGCAAGCTGACGCGCTGGAGCGCATCCGC 60
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGluLysIleArg 20
QY 61 CTGCGCGCCGCGCGCAAGAGTGCTACATGATGAAGCACTGTGTGGCCAGCGCGCAG 120
DB 21 LeuArgProGlyGlyLysLysArgTyrMetIleLysHisLeuValTrpAlaSerArgGlu 40
QY 121 CTGAGAAGTTCGCTGAAACCCGCGCTGTGAGACCAAGCGGCTGCAAGAGATC 180
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuGluThrSerGlyCysLysGlnIle 60
QY 181 ATCCGCGAGCTGCACCCCGCTGCAACCGCGAGCGAGAGCTGAAGACCTGTTCAC 240
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGTGACAGAGAAGATCGAGCTCCGCGACCAAGAGAGCC 300
DB 81 ThrValAlaThrLeuTyrCysValHisGluGlyValGluValArgAspThrLysGluAla 100
QY 301 CTGACAAGATCGAGAGAGAGAGACAAGTGCCAGAGAGATCCAGCAG-----GCC 354
DB 101 LeuAspArgIleGluGluGluGlnAsnLysIleGlnGlnLysIleGlnGlnLysThrGln 120
QY 355 GAGCGCGCCGCAAGGCGCAAGTGAGCCAGACTACCCCATCGTGAGAGAGCTGAGGGC 414
DB 121 GlnAlaAlaAsp--GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 139
QY 415 CAGATGCTGCACCAAGGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGTGATGAG 474
DB 140 GlnMetValHisGlnLysLeuSerProArgThrLeuAsnAlaTrpValLysValIleGlu 159
QY 475 GAGAAAGCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCAC 534
DB 160 GluLysAlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThr 179
QY 535 CCCCAGGACCTGAACAGATGTGAACACCGTGGCGGCCACCAAGCGCCCATGACATG 594
DB 180 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMet 199
QY 595 CTGAAGGACACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCC 654
DB 200 LeuLysAspThrIleAsnGluGlnAlaAlaGluTrpAspArgLeuHisProValHisAla 219
QY 655 GGGCCCATCGCCCCCGGCGAGATGCGCGAGCGCCGCGCGAGCATCGCCGCGACCAAC 714
DB 220 GlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 239
QY 715 AGCACCTTGACGAGAGAGATCGCTGTGATGACAGCAACCCCGCATCCCGTGGGCGAC 774
DB 240 SerThrLeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAsp 259
QY 775 ATCTACAAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTG 834
  
```

```

DB 260 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProVal 279
QY 835 AGCATCCTTGACATCAAGCAGGCGCCCAAGAGAGCCCTCCGCACTACGTGAGCGCTTC 894
DB 280 SerIleLeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 299
QY 895 TTCAAGACCTCGCGCGCGAGCAGAGCAACCAAGAGTGAAGAACTGATGACCGACAC 954
DB 300 PheLysThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThr 319
QY 955 CTGCTGTGACAGAACGCCAACCCGCACTGCAAGACCATCTCGCGCTCTCGCGCCGCC 1014
DB 320 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGly 339
QY 1015 GCCAGCTGAGAGATGATGACCGCTGCAAGGCGCTGGCGCGCCAGCCAGAGGCC 1074
DB 340 AlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyProSerHisLysAla 359
QY 1075 CGCGTGTGCGCGCGAGCGATGAGCCAGGCCAACAC--AGCGTATGATGACAGAGAGC 1131
DB 360 ArgValLeuAlaGluAlaMetSerGlnThrAsnSerValAsnIleLeuMetGlnLysSer 379
QY 1132 AACTTCAAGGCGCGCGCATGCTCAAGTGTCTTCAACTGCGGCAAGAGGCGCCATC 1191
DB 380 AsnPheLysGlyAsnLysArgMetValLysCysPheAsnCysGlyLysGluGlyHisIle 399
QY 1192 GCGCGCACTGCGCGCGCGCGCGCGCGCAAGAGGCTGTGAAAGTGGCGCAAGAGGCGCAC 1251
DB 400 AlaArgAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGluGlyHis 419
QY 1252 CAGATGAAGACTGCAACCGAGCGCCAGCCACTTCTGGGCAAGATCTGGCCAGCCAC 1311
DB 420 GlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHis 439
QY 1312 AAGGCGCGCGCGCGCACTTCTTGCAGAGAGCGCGCGCGAGCCCGCGCGCGAG 1371
DB 440 LysGlyArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGlu 459
QY 1372 AGCTTCCGCTTGCAGAGAGACCAACCCCGCGCAGAAAGAGAGAGAGAGAGAGAGC 1431
DB 460 SerPheArgPheGluGluThrThrProAlaProLysGlnGlnLysAspArgGluAla 479
QY 1432 CTGACCAGCTGAAGAGCCTGTTCGCGCAACGAGCCCGCTGAGCCAG 1476
DB 480 LeuThrSerLeuLysSerLeuPheGlySerAspProLeuSerGln 494
  
```

RESULT 12

```

AD C72910
ID AD C72910 standard; protein; 491 AA.
XX
AC AD C72910;
XX
DT 18-DEC-2003 (first entry)
XX
DE HIV-1 gag protein containing an NCP sequence, SEQ ID No 140.
XX
KW complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;
KW human immunodeficiency virus; HIV.
XX
OS Human immunodeficiency virus 1.
XX
PN WO2003060098-A2.
XX
PD 24-JUL-2003.
XX
PF 10-JAN-2003; 2003WO-US000801.
XX
PR 11-JAN-2002; 2002US-0347369P.
XX
PA (ACHI-) ACHILLION PHARM INC.
XX
PI Beuchter D, Hou X, Marlor CW, Rice WG, Yang W;
  
```

XX WPI; 2003-646042/61.  
DR  
XX Method of determining whether a compound inhibits formation of complex  
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
PT oligonucleotide by comparing amount of complex formed in presence/absence  
of compound.  
XX  
PS Claim 2; SEQ ID NO 140; 105bp; English.  
XX  
CC The invention relates to a novel method of determining whether a compound  
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (Ncp7)  
CC polypeptide and an HIV-psi-site oligonucleotide. The method involves  
CC adding the HIV-psi-site oligo to a mixture of the Ncp7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an Ncp sequence of the invention.  
XX  
SQ Sequence 491 AA:  
  
Alignment Scores:  
Pred. No.: 9.12e-136 Length: 491  
Score: 2415.50 Matches: 450  
Percent Similarity: 96.75% Conservative: 26  
Best Local Similarity: 91.46% Mismatches: 15  
Query Match: 87.14% Indels: 1  
DB: 7 Gaps: 1  
  
US-09-475-704A-3 (1-1479) x ADC72910 (1-491)  
QY 1 ATGGGGCGCCCGCCGACGATCCTCGCGCGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCCCGCGGCAAGAAGTGTATCATGATGAAGCACTGTGTGGCCAGCGCGGAG 120  
DB 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGAAGTTCGCTTGAACCCCGGCTGTGAGAACCGAGGGGCTGACAGCATC 180  
DB 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60  
QY 181 ATCCGCCAGCTGCACCCCGCCCTGTGAGACCGGACGAGAGCTGAAGCCTGTTCAC 240  
DB 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTCTGTACTGCGTGCACGAGAAGTCGAGTCCCGGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisAlaGluIleGluValArgAspThrLysGlnAla 100  
QY 301 CTGACAAGATCGAGAGAGGACGACAAAGTCCAGACGAAAGATCCAGAGCGCGAGGCC 360  
DB 101 LeuAspArgIleGluGluGlnGlnAsnLysSerGlnGlnLysThrGlnAlaAsnGlu 120  
QY 361 GCCGACAAGGGCAAGGTGAGCCGAACTACCCCATCTGTGAGAACTTGACGGGCCAGATG 420  
DB 121 AlaAsp--GlyLysValSerGlnAsnLysProIleValGlnAsnLeuGlnGlyMet 139  
QY 421 GTGCACACGAGCCATCAGCCCCCGGACCTGAACGCGCTGGTGAAGGTGATCGAGAGAAG 480  
DB 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLys 159  
QY 481 GCCTTCAGCCCGGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCCACCCCCAG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179  
QY 541 GACCTGAACACGATGTTGAACACCGGTGGCGGCGCACGAGCGCCGATGACGATGCTGAAG 600

DB 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAAGAGAGGCGCGGAGTGGGACCGCGTGACACCCCGTGACCGCGGCCCC 660  
DB 200 AspThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 219  
QY 661 ATCGCCCCCGGCGAGATGCGGAGAGCCCCCGGCGACGACATGCGCGGACCAAGCACC 720  
DB 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerSer 239  
QY 721 CTGCAGAGAGATCGCCTGTGATGACCAAGCAACCCCCCATCCCGTGGGCGCATCTAC 780  
DB 240 LeuGlnGlnGlnIleAlaTrpMetThrGlyAsnProProValProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 840  
DB 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279  
QY 841 CTGACATCAAGCAGGGCGCCCAAGAGAGCCCTTCGCGACTACGTGACCGCTTCTCAAG 900  
DB 280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGGACGACAGCAACCCAGAGTGAAGAACTGATGACCGACCCCTGCTG 960  
DB 300 ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGGCGCGCGCGCAGC 1020  
DB 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 339  
QY 1021 CTGAGAGAAGTATGACCGCGCTGCGAGGGCGGTGGCGCGCCAGCCACMAAGCCGCGTG 1080  
DB 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359  
QY 1081 CTGGCGGAGGCGATGAGCCAGGCCAACCAACGCGTGATGATGCAGAAGACAACTTCAAG 1140  
DB 360 LeuAlaGluAlaMetSerGlnThrAsnSerThrIleLeuMetGlnArgSerAsnPheLys 379  
QY 1141 GAGCCCGCGCGCATCTGTCAGTGTCTTCAACTGCGGCAAGAGGGCCACATGCGCGCAAC 1200  
DB 380 GlyProLysArgIleValLysCysPheAsnGlyLysGluGlnHisIleAlaLysAsn 399  
QY 1201 TGCCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACGATGAAG 1260  
DB 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisGlnMetLys 419  
QY 1261 GACTGCACCGAGCGCGGCAAGCTTCTGCGCAAGATCTGAGCCAGCAAGAGGGCGCGC 1320  
DB 420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCGCGCAACTTCTGACAGAGCGCGCGGAGCCACCGCGCGCGCGCGGAGACTTCCGC 1380  
DB 440 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 459  
QY 1381 TTCGAGAGACCAACCCCGCGCGCAAGAGAGAGCAAGGACCGGAGACCTGACCCAGC 1440  
DB 460 PheGluGluThrThrProAlaLeuGlnGlnGlyProLysAspArgGluProLeuThrSer 479  
QY 1441 CTGAAGAGCCTGTTGCGCAAGACCCCGTGAAGCAG 1476  
DB 480 LeuArgSerLeuPheGlySerAspProLeuSerGln 491  
  
RESULT 13  
ID ADC72874  
ID ADC72874 standard; protein; 491 AA.  
XX  
AC ADC72874;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE HIV-1 gag protein containing an Ncp sequence, SEQ ID No 104.  
XX complex; HIV nucleocapsid protein 7; Ncp7; HIV-psi-site; anti-HIV;

KW human immunodeficiency virus; HIV.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO2003060098-A2.  
XX  
PD 24-JUL-2003.  
XX  
PF 10-JAN-2003; 2003WO-US000801.  
XX  
PR 11-JAN-2002; 2002US-0347369P.  
XX  
PA (ACHI-) ACHILLION PHARM INC.  
XX  
PI Beuchter D, Hou X, Marlor CW, Rice WG, Yang W;  
XX  
DR WPI; 2003-646042/61.  
XX  
PT Method of determining whether a compound inhibits formation of complex  
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
PT oligonucleotide by comparing amount of complex formed in presence/absence  
PT of compound.  
XX  
PS Claim 2; SEQ ID NO 104; 105pp; English.  
XX  
CC The invention relates to a novel method of determining whether a compound  
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (Ncp7)  
CC polypeptide and an HIV-psi-site oligonucleotide. The method involves  
CC adding the HIV-psi-site oligo to a mixture of the Ncp7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an Ncp sequence of the invention.  
XX  
SQ Sequence 491 AA;  
  
Alignment Scores:  
Pred. No.: 2.73e-135 Length: 491  
Score: 2407.50 Matches: 449  
Percent Similarity: 96.34% Conservative: 25  
Best Local Similarity: 91.26% Mismatches: 17  
Query Match: 86.85% Indels: 1  
DB: 7 Gaps: 1  
  
US-09-475-704A-3 (1-1479) x ADC72874 (1-491)  
QY 1 ATGGGGCGCCGCGCGCATCTGCGGGCGGCAAGCTGACGCGCTGGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuAGlyGluLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCCCGCGCGCAAGAGTGTATGATGTAAGCACTGTGTGGCCAGCCGCGAG 120  
DB 21 LeuArgProGlyGlyArgLysHisTrpMetLeuLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGAAGTTCGCGCTGAACCCCGGCTGTGAGACCAAGGAGCTGCAAGCATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysLysGlnIle 60  
QY 181 ATCGGCCAGCTGACACCCCGCCCTGCAACCGGCAAGAGTCAAGCTGTCAAC 240  
DB 61 IleLysGlnLeuHisProAlaLeuLysThrGlyThrGluLysLeuArgSerLeuLysAsn 80  
QY 241 ACCGTGGCCACCCCTTACTGCGTGCACAGAGAAGATCGAGTCCGCGACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuLysCysValHisGluAsnIleGluValArgAspThrLysGluAla 100  
QY 301 CTGACACAAGATCGAGAGAGAGCAAGACAAGTGCACAGAGAAGATCCAGCAGCGCGGCC 360  
DB 101 LeuAspLysIleGluGluGluGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120

QY 361 GCCGACAAGGCAAGGTGAGCCAGAACTACCCCATGCTGCAAGAACTGCAAGGCCAGATG 420  
DB 121 AlaAspGluGly--ValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyGlnMet 139  
QY 421 GTGACACAGGCCATGACCCCGCGACCCCTGAACGCGCTGGGTGAAGGTGATCGAGAGAAG 480  
DB 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159  
QY 481 GCCTTACAGCCCGGAGGTGATCCCACTGTTCACCGCCCTGAGCGAGGGGCCACCCCGCAG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGluAlaThrProGln 179  
QY 541 GACCTGAACAGATGTTGAACACCCGTGGCGCGGCCACAGCCCGCATGCAAGATGCTGAAG 600  
DB 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAACGAGAGAGCGCCGCGAGTGGACCGCGTGCACACCCCGTGCACCGCGGCC 660  
DB 200 AspThrIleAsnGluGluAlaIleGluTrpAspArgLysHisProValHisAlaGlyPro 219  
QY 661 ATCGCCCGCGCGCAAGATGCGCGAGCGCCGCGGCGAGCAATCGCCGCGCACCAAGCACCC 720  
DB 220 AlaAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGCAAGAGCAGATGCGCTGTGATGACCAAGCAACCCCGCATCCCGCTGGCGGCGACATAC 780  
DB 240 LeuGlnGluGlnIleAlaTrpMetThrGlyAsnProProValProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGGCGCTGAACAAGATCGTGGCGATGTACAGCCCGTGAGCATC 840  
DB 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279  
QY 841 CTGACATCAAGCAGAGGCGCCCAAGAGCCCTTCCCGACTACGTGACCGCTTCTTAAG 900  
DB 280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGCGAGCAGACCAAGAGGTGAAGAAGTGAAGTGAACCGACACCCCTGCTG 960  
DB 300 ValLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCGCGCGCGCAC 1020  
DB 320 IleGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAlaIleSer 339  
QY 1021 CTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCGACCAAGGCGCGGTG 1080  
DB 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgVal 359  
QY 1081 CTGGCCGAGCGATGAGCCAGGCCAACCAACGCGTGTATGATGCAAGAAGCAACTTCAAG 1140  
DB 360 LeuAlaGluAlaMetSerGlnAlaAsnSerAsnIleMetMetGlnArgSerAsnPheLys 379  
QY 1141 GGCCCGCGCGCATGCTCAAGTGTCTCACTGCGCAAGAGGCGCATCGCCCGCAAC 1200  
DB 380 GlySerLysArgIleValLysCysPheAsnCyseGlyLysGluGlnHisIleAlaArgAsn 399  
QY 1201 TGCCGCGCGCCCGCAAGAGGCTGTGAAGTGGCGGCAAGAGGCGCCACCAAGTGAAG 1260  
DB 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlnHisIleGlnMetLys 419  
QY 1261 GACTGACCGAGCGCGAGGCCAACTTCTGGGCAAGATCTGGCCAGCCACCAAGGCGCC 1320  
DB 420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCCGCAACTTCTGACAGAGCGCGCCGAGCCACCGCCCGCGCGAGAGCTTCGCG 1380  
DB 440 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 459  
QY 1381 TTCGAGAGAGACCAACCCCGCGCAAGAGCAGAGCAAGAGCAGACCCGAGACCTGACAGC 1440  
DB 460 PheGluGluThrThrProAlaProLysGlnGlnSerLysAspArgGluProLeuIleSer 479

QY 1441 CTGAAGACCTGTTGGCAACGACCCCTGAGCCAG 1476  
Db 480 LeuYsSerLeuPheGlySerAspProSerSerGln 491

RESULT 14

AAB69278  
ID AAB69278 standard; protein; 491 AA.  
XX

AC AAB69278;

DT 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

XX HIV-1 non-subtype B clone 94IN476-104 gag protein.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpr; vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

XX WO200026416-A1.

PD 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837.

XX 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

XX Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Claim 41; Fig 14; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX SQ Sequence 491 AA;

Alignment Scores: Pred. No.: 3.6e-135 Length: 491

Score: 2405.50 Matches: 449

Percent Similarity: 96.54% Conservative: 26

Best Local Similarity: 91.26% Mismatches: 16

Query Match: 86.78% Indels: 1

DB: 3 Gaps: 1

US-09-475-704A-3 (1-1479) x AAB69278 (1-491)

QY 1 ATGGGCGCGCGCGCAAGCATCTGCGCGCGCGCAAGCTGAGCGCCTGGAGCGCATCCGC 60

Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspArgTrpGluLysIleArg 20

QY 61 CTGCGCGCGCGCGCAAGAGTGTATCATGATGAAGCACTGTGTGGGCGCAGCGCGAG 120

Db 21 LeuArgProGlyGlyLysLysHisTyrMetIleLysHisIleValIleTyrAlaSerArgGlu 40

QY 121 CTGGAAGAAGTTCGCGCTGAACCCCGCGCTGTGAGAGACCGAGGCGCTGCAAGCAGATC 180

Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerArgGlyCysLysGlnIle 60

QY 181 ATCCGCCAGCTGCAACCCCGCTGCAAGCCGGCAGCGAGAGCTGAAGAGCTGTTCAC 240

Db 61 IleYsGlnLeuHisProAlaLeuLysThrGlyThrGluLeuArgSerLeuPheAsn 80

QY 241 ACCGTGGCCACCTGTACTGCGCTGCAAGAGATCGAGGTCCGCGACCAAGAGAGCC 300

Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluValArgAspThrLysGluAla 100

QY 301 CTGGAACAAGATGAGAGAGAGCAACAAGTGCAGAGAGATCCAGAGGCGCGAGCC 360

Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerGlnGlnIleGlnGlnAlaLysGlu 120

QY 361 GCCGACAAGGCGAAGGTGAGCCGAACCTACCCCATCGTGCAAGACCTGACGGCGCAGATG 420

Db 121 AlaAsp--GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 139

QY 421 GTGCACCAAGGCGCATCAGCCCGCGCACCTGTGAAGCGCTGGTGAAGTGATCGAGAGAG 480

Db 140 ValHisGlnProLeuSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159

QY 481 GCCTTCAGCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGGCGCGCACCCCGAG 540

Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProSer 179

QY 541 GACCTGAACAGATGTTGAACAACGTGGCGCGCACAGGCGCGCATGACAGTGTGAAG 600

Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLys 199

QY 601 GACACCATCAAGAGAGGCGCGCGAGTGGGACCGCGTGACCCCGTGACCGCGCGCCC 660

Db 200 AspThrIleAsnGluGluValAlaIleGluTrpAspArgLeuHisProValHisAlaGlyPro 219

QY 661 ATCGCCCGCGCGAGATGCGCGAGCCCGCGCGAGCAGACATGCGCGCACCAAGCAGACC 720

Db 220 AsnProProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239

QY 721 CTGCAGAGAGATGCGCTGTGAAGCAACCCCGCATCCCGTGCGCGCATCTAC 780

Db 240 LeuGlnGluGlnIleAlaTrpMetThrGlyAsnProProIleProValGlyAspIleTyr 259

QY 781 AAGCGTGATCATCTGCGCTGAACAAGATCGTGCAGATGACGCCCGTGAGCATC 840

Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279

QY 841 CTGACATCAAGCAGGCGCGCGAGGCGCGCTTCGCGACTGACGTGACCGCTTCTTCAAG 900

Db 280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299

QY 901 ACCCTGCGCGCGAGCAGAGCAACCGAGGTGAAGAACTGATGACCGACACCTGCTG 960

Db 300 ThrLeuArgAlaGluGlnAlaThrGlnGluValLysGlyTyrMetThrAspThrLeuLeu 319

QY 961 GTGCAGAACGCCAACCCCGACTGCAAGACCACTCTGCGCGCTCTGCGCGCGCGCGCAGC 1020

Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 339

QY 1021 CTGAGAGAGATGATGACCGCTGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCTG 1080

Db 340 LeuGluGluMetValThrAlaCysGlnGlyValGlyGlyProSerHisLysValAlaArgVal 359

QY 1081 CTGGCCGAGCGGATGAGCGCAAGCCCAACACAGCGTGTGATGACAGAGCACTTCAAG 1140

Db 360 LeuAlaGluAlaMetSerGlnSerHisSerAsnIleMetMetGlnArgGlyAsnPheLys 379

QY 1141 GCGCCCGCGCGCATCGTCAAGTGTCTCAACTGCGCGCAAGGAGGCGCACATGCGCGCAAC 1200

Db 380 GlyProLysArgGluValLysCysPheAsnCysGlyLysGluGlnHisIleAlaArgAsn 399

QY 1201 TGCCGCGCGCGCGCAGAGAGGCGTGTGAAGTGCAGAGAGGCGCGCACAGATGAAG 1260

Db 400 CysArgAlaProArgLysArgGlyCysTrpLysCysGlyGlnGluGlnHisGlnMetLys 419

QY 1261 GACTGCACCGAGCGCGCAAGCCCACTTCTGCGCAAGATCTGCGCGCACGCAAGAGGCGCGC 1320

```
Db      420 AspCySThrGluArgInAlaAsnPhenLeuGlyLysIleTrpProSerHisLysGlyArg 439
|||
QY      1321 CCGGCACACTTCCTGCAGAGCCGCCGAGCCGCCGCCGCCGCCGCCGAGAGCTTCCGC 1380
|||||
Db      440 ProGlyAsnPhenLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPheArg 459
|||||
QY      1381 TTCGAGAGAGACCACCCCCGGCACAAGACGAGAGACCAAGACCGCGAGACCTGACCAGC 1440
|||||
Db      460 PheLysGluThrThrProAlaProLysGlnGluSerLysAspArgGluProLeuThrSer 479
|||||
QY      1441 CTGAAGAGCCTGTTCGCAACGACCCCTTGAGCCAG 1476
|||||
Db      480 LeuLysSerLeuPheGlySerAspProLeuSerGln 491
|||||

RESULT 15
ADCT2913
ID     ADC72913 standard; protein; 491 AA.
XX
AC     ADC72913;
XX
DT     18-DEC-2003 (first entry)
XX
DE     HIV-1 gag protein containing an Ncp sequence, SEQ ID No 143.
XX
KW     complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;
KM     human immunodeficiency virus; HIV.
XX
OS     Human immunodeficiency virus 1.
PN     WO2003060098-A2.
PD     24-JUL-2003.
PF     10-JAN-2003; 2003WO-US000801.
PR     11-JAN-2002; 2002US-0347369P.
PA     (ACHI-) ACHILLION PHARM INC.
PI     Beauchter D, Hou X, Marlor CW, Rice WG, Yang W;
XX
DR     WPI; 2003-646042/61.
XX
PT     Method of determining whether a compound inhibits formation of complex
PT     between HIV nucleocapsid protein 7 polypeptide and HIV psi-site
PT     oligonucleotide by comparing amount of complex formed in presence/absence
PT     of compound.
XX
PS     Claim 2; SEQ ID NO 143; 105pp; English.
XX
CC     The invention relates to a novel method of determining whether a compound
CC     inhibits formation of a complex between HIV nucleocapsid protein 7 (Ncp7)
CC     polypeptide and an HIV-psi-site oligonucleotide. The method involves
CC     adding the HIV-psi-site oligo to a mixture of the Ncp7 polypeptide and
CC     the novel compound and comparing amount of complex formed, with that
CC     formed in the absence of the compound, where a decrease in the amount of
CC     complex formed in presence of the compound indicates that the compound
CC     inhibits complex formation. The method of the invention involves anti-HIV
CC     activity. The compounds identified are useful for treating a subject
CC     infected with human immunodeficiency virus (HIV) by administering the
CC     compound to the subject. This sequence represents an HIV-1 gag protein
CC     which contains an Ncp sequence of the invention.
XX
SQ     Sequence 491 AA;

Alignment Scores:
Pred. No.:          1.24e-134           Length:         491
Score:              2396.50             Matches:          449
Percent Similarity: 95.53%               Conservative:     21
Best Local Similarity: 91.26%            Mismatches:       21
Query Match:        86.45%               Indels:           1
DB:                  7                    Gaps:             1
```

US-09-475-704A-3 (1-1479) x ADC72913 (1-491)
QY 1 ATGGGCGCGCGCCAGCATCTGTGCGGCGGCAAGCTGGAGCGCTGGAGCGCATCCGC 60
Db 1 MetG1yAlaArgAlaSerIleLeuArgG1yG1yLysLeuAspLysTrpG1uLysIleArg 20
QY 61 CTGGCGCGCGCGGCAAGAAGTGTCTACATGATGAAGCACTGTGTGTGGGCCAGCCGGAG 120
Db 21 LeuArgProG1yG1yLysLysHisIstYrMetIleLysHisLeuValTrpAlaSerArgGlu 40
QY 121 CTGGAGAAAGTTGCGCCCTGAACCCCGGCTGTCTGGAGAGCCAGCGAGGGCTGCAAGCATC 180
Db 41 LeuGluArgPheAlaLeuAsnProG1yLeuLeuGluThrSerG1uG1yCysLysGlnIle 60
QY 181 ATCCGCCAGCTGCACCCCGCGCTGCAGACCGGCGAGCGAGAGCTGAAGAGCCTGTTCAAC 240
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrG1yThrG1uG1uLeuArgSerIleuHisAsn 80
QY 241 ACCGTGGCCACCCCTGTACTGCGCTGCACGAGAAAGATCGAGGTCGCCGACACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisAlaG1yIleG1uIleArgAspThrLysGluAla 100
QY 301 CTGGACAAAGATGAGGAGGAGAGCAACAAGTGCACGAGAGATCCAGCAGCGCGAGGCC 360
Db 101 LeuAspLysIleG1uG1uG1uG1uLysSerGlnG1uLysThrGlnAlaLysGlu 120
QY 361 GCCGACAAAGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAAACTGCAGGGCCAGATG 420
Db 121 AlaAsp--G1yLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnG1yGlnMet 139
QY 421 GTGCACCAAGGCATCAGCCCGCGCACCCCTGAACGCCCTGGGTGAAGGTGATCGAGGAGAG 480
Db 140 ValHisGlnAlaLeuSerProArgThrLeuAsnAlaTrpValLysValIleG1uG1uLys 159
QY 481 GCCTTCAGCGCGCGAGGTGATCCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540
Db 160 AlaPheSerProG1uIleIleProMetPheThrAlaLeuSerG1uG1yAlaThrProGln 179
QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGGACCAAGCGCCCATGACAGATGCTGAAG 600
Db 180 AspleuAsnThrMetLeuAsnThrValG1yG1yHisGlnAlaAlaMetGlnMetLeuLys 199
QY 601 GACACCATCAACGAGGAGGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGGCCCC 660
Db 200 AspThrIleAsnG1uG1uAlaAlaG1uTrpAspArgLeuHisProAlaGlnAlaG1yPro 219
QY 661 ATGCCCCCGGCGAGATGCGCGAGCGCGCGGCGAGGACATCGCGCGGACCAACGAGACC 720
Db 220 IleAlaProG1yGlnMetArgGluProArgG1ySerAspIleAlaG1yThrThrSerThr 239
QY 721 CTGCAGAGAGCATCGCCTGGATGACCAAGCAACCCCGCCATCCCGTGGCGGACATCTAC 780
Db 240 LeuGlnG1uGlnIleAlaTrpMetThrG1yAsnProProValProValG1yGlnIleTyr 259
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGCGGATGTACAGCCCGGTGAGCATC 840
Db 260 LysArgTrpIleIleLeuG1yLeuAsnLysIleValArgMetTyrSerProValSerIle 279
QY 841 CTGGACATCAAGCAGGGGCGGCAAGAGCGCCTTCCGCGACTACGTGAGCCGCTTCTCAAG 900
Db 280 LeuAspIleLysGlnG1yProLysGluProPheArgAspTyrValAspArgPhePheLys 299
QY 901 ACCCTGCGCGCGGAGCAGAGCAACCCAGGAGGTGAAGAACTGAGATGACCGGACCCCTGCTG 960
Db 300 ThrLeuArgAlaG1uGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319
QY 961 GTGCAGAAACGCCCAACCCCGACTGCAGAGCAATCTGCGCGCTCTGGCGCGCGCGGAGC 1020
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuG1yProG1yAlaSer 339
QY 1021 CTGGAGAGATGATGACCGCGCTGCCAGGGCGTGGGGCGGCGCCAGCCAGAGCGCGCGTG 1080

```
Db      340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359
QY      1081 CTGGCCGAGGGCGATGAGCCAGGCCAACACCAACCGTATGATGACAGAAGAGCAACTTCAAG 1140
Db      360 LeuAlaGluAlaMetSerGlnThrAsnAsnSerIleLeuMetGlnArgSerAsnPheLys 379
QY      1141 GGGCCCCGGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATGCGCCGCAAC 1200
Db      380 GlyPheLysArgThrValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 399
QY      1201 TGGCGGCGCCCCCGCAAGAAAGGGCTGTGGAAGTGGGCAAGGAGGGCCACAGATGAAG 1260
Db      400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 419
QY      1261 GACTGCACCGAGCGCCAGGGCCAACTTCTGGGCAAGATCTGGCCAGCCACAAGGGCCGC 1320
Db      420 AspCysThrGluArgGlnAlaAsnPheLeuGlySileTrpProSerHisLysGlyArg 439
QY      1321 CCGGGCAACTTCTGTGACAGAGCCGGCCGAGCCCAACCGCCCCCGCGAGAGCTTCCGC 1380
Db      440 ProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPheArg 459
QY      1381 TTGAGAGAGACCAACCCCGGCGCAGAGCAGAGAGCAAGACCGCGAGACCTGACCAGC 1440
Db      460 PheGluGluThrThrProAlaLeuLysGlnGluLysAspArgGluProLeuThrSer 479
QY      1441 CTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAG 1476
Db      480 LeuLysSerLeuPheGlySerAspProLeuSerGln 491
```

Search completed: March 11, 2005, 15:22:11  
Job time : 242.694 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:04:55 ; Search time 51.7254 Seconds  
(without alignments)  
5502.310 Million cell updates/sec

Title: US-09-475-704A-3  
Perfect score: 2772  
Sequence: 1 atggcgccgcgcgcacgcat.....acgacccctgagccagctaa 1479

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+n2p.model -DEV=xlp  
-Q=/cg2.1/USPTO\_spool\_p/US09475704/runat\_10032005\_140222\_14839/app\_query.fasta\_1.3342

-DB=PIR\_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09475704 @CGN\_1\_1\_160 @runat\_10032005\_140222\_14839 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2300.5	83.0	501	2	S54377	gag polypotein -
2	2294.5	82.8	497	1	FOLJND	gag polypotein -
3	2266.5	81.8	505	2	T01667	gag polypotein -
4	2263	81.6	500	1	FOVWL	gag polypotein -
5	2249	81.1	500	1	FOVWH4	gag polypotein -
6	2247	81.1	502	1	FOVMA2	gag polypotein -
7	2235	80.6	512	1	FOVWH3	gag polypotein -
8	2233	80.6	498	2	T09436	gag polypotein -
9	2231	80.5	500	1	A44001	gag polypotein -
10	2228	80.4	500	2	S33979	gag polypotein -
11	2216.5	80.0	506	1	A38068	gag polypotein -
12	2037.5	73.5	478	1	FOVWL	gag polypotein -
13	1989	71.8	508	1	FOLJSI	gag polypotein -
14	1400	50.5	521	2	S08435	gag polypotein -

15	1399.5	50.5	521	2	S53091	gag polypotein -
16	1398	50.4	521	1	FOLJST	gag polypotein -
17	1392.5	50.2	554	2	S46346	gag polypotein -
18	1380	49.8	521	1	FOLJCA	gag polypotein -
19	1376	49.6	522	1	FOLJG6	gag polypotein -
20	1375	49.6	521	2	S12152	gag polypotein -
21	1371	49.5	507	2	T11559	gag polypotein -
22	1370	49.4	507	2	S04237	gag polypotein -
23	1364.5	49.2	522	1	FOLJG2	gag polypotein -
24	1364	49.2	510	1	FOLJTM	gag polypotein -
25	1354.5	48.9	506	1	FOLJG3	gag polypotein -
26	1354	48.8	519	1	FOLJG4	gag polypotein -
27	1341.5	48.4	506	1	FOLJG5	gag polypotein -
28	1222.5	44.1	502	2	S28080	gag polypotein -
29	716.5	25.8	171	2	S52929	GAG protein - huma
30	692	25.0	146	2	S60708	gag protein - huma
31	688	24.8	146	2	S60698	gag protein - huma
32	688	24.8	146	2	S60704	gag protein - huma
33	688	24.8	146	2	S60697	gag protein - huma
34	683	24.6	146	2	S60699	gag protein - huma
35	683	24.6	146	2	S60702	gag protein - huma
36	680	24.5	146	2	S60700	gag protein - huma
37	680	24.5	146	2	S60703	gag protein - huma
38	560	20.2	212	2	S03070	gag polypotein -
39	454.5	16.4	450	1	FOLJFP	gag polypotein -
40	453.5	16.4	450	2	S23819	gag protein - feli
41	449	16.2	449	2	A45557	matrix, capsid, nu
42	445	16.1	486	1	FOLJBY	gag polypotein -
43	441	15.9	85	2	S49086	gag protein - huma
44	435.5	15.7	450	2	S25162	gag protein - feli
45	431.5	15.6	1106	2	JQ0405	hypothetical 119.5

ALIGNMENTS

RESULT 1  
S54377  
gag polypotein - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004  
C:Accession: S54377  
R:Theodore, T.; Buckler-White, A.J.  
submitted to the EMBL Data Library, July 1989  
A:Reference number: S54377  
A:Accession: S54377  
A:Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-501 <THR>  
A:Cross-references: UNIPROT:P12495; EMBL:M22639; NID:g329377; PIDN:AAA45365.1; PID:g3293  
C:Superfamily: AIDS-related virus gag polypotein  
C:Keywords: polypotein

Alignment Scores:

Pred. No.: 1.03e-109 Length: 501  
Score: 2300.50 Matches: 434  
Percent Similarity: 93.21% Conservative: 33  
Best Local Similarity: 86.63% Mismatches: 25  
Query Match: 82.99% Indels: 9  
DB: 2 Gaps: 4

US-09-475-704A-3 (1-1479) x S54377 (1-501)

QY	1	ATGGCGCCCGCGCCAGCATCTCGCGCGCGCAAGCTGACGCGCTGGAGCGCATCCGC	60
DB	1	MetGlyAlaArgIaSerValLeuSerGlyGlybLeuAspAlaTrpGluIleArg	20
QY	61	CTGGCGCCCGCGCGCAAGAGTCTACATGATGAACACCTGTGTGGCCAGCGCGAG	120
DB	21	LeuArgProGlyGlybLeuIleTyrArgLeuIleValTrpAlaSerArgGlu	40
QY	121	CTGAGAAGTTTCGCTGACCCCGCTGCTGAGACACGAGCGGCTGCAAGCATC	180
DB	41	LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerAspGlyCysIleGlnIle	60

```
OY 181 ATCCGCCAGCTGCACCCCGCCCTGCAGACCGGCGAGCGAGAGCTGAAGACCTGTTCAC 240
    |||||
Db 61 ILeGlyInLeuGlnProAlaIleArgThrGlySerGluLeuArgSerLeuPheAsn 80
OY 241 ACCGTGGCCACCTGTACTGCGGTGCACGAGAGATCGAGTCCGCGACACCAAGAGGCC 300
    |||||
Db 81 ThrValAlaThrLeuYrCySvalHisGluArgIleGluValLysAspThrLysGluAla 100
OY 301 CTGAGCAAGATCGAGAGAGAGAGCAAGATGCGACGAGAGATCCAGAGCGCGAGGCC 360
    |||||
Db 101 LeuGluYsMetGluGluGluGlnAsnLysSerLysAsnLysLysAlaGlnGlnAlaAla 120
OY 361 GCCGCAAGGGC-----AAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAG 411
    |||||
Db 121 AlaAspAlaGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGln 140
OY 412 GGCCAGATGGTGACACCGCCCATCAGCCCGCCGACCCCTGAAAGCGCTGGAGGTGATC 471
    |||||
Db 141 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaIleValLysValIle 160
OY 472 GAGGAGAGGGCCTTCAGCCCGGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGGCC 531
    |||||
Db 161 GluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 180
OY 532 ACCCCCGAGGACCTGAACACGATGTTGAACACCGTGGCGGGCCACCGAGCGCCCATGCA 591
    |||||
Db 181 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 200
OY 592 ATGTGTAAGGACACCATCAACGAGAGGGCGCGGAGTGGGACCGCGTGCACCCCGTGCAC 651
    |||||
Db 201 MetLeuLysGluThrIleAsnGluGluAlaAlaGluIleTrpAspArgLeuHisProValHis 220
OY 652 GCCGGCCCATCGCCCGCGGCGAGATGCGCGAGCGCGCGGCGAGCGACATCGCCGCGACC 711
    |||||
Db 221 AlaGlyProIleAlaProGluGlnMetArgGluProArgGlySerAspIleAlaGlyThr 240
OY 712 ACCAGCACCTGCAGAGAGATCGCTGGATGACGACCAACCCCCCATCCCGCTGGCGC 771
    |||||
Db 241 ThrSerThrLeuGlnGluGlnIleAlaIleTrpMetThrSerAsnProIleProValGly 260
OY 772 GACATCTCAAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGCGATGTACAGCCC 831
    |||||
Db 261 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 280
OY 832 GTGAGCATCTGCAGATCAAGCAGGGCGCCCAAGGCCCTTCCGCGACTACGTGACCGC 891
    |||||
Db 281 ValSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 300
OY 892 TTCTTCAAGACCTCGCGCGCGCGAGCAGACCAACCGAGAGGTGAAGAACTGATGACCGAC 951
    |||||
Db 301 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysGlyTyrMetThrGlu 320
OY 952 ACCCTGCTGTGAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCGCTTCGCGCCC 1011
    |||||
Db 321 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 340
OY 1012 GCGCGCACCTGCAGAGAGATGATACCGCTGCGAGGGCGGTGGGCGGCCCGACCAAG 1071
    |||||
Db 341 GlnAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLys 360
OY 1072 GCCCGCGTGTGCGCAGGCGGATGACCGCAAGCCCAACCAAGC-----GTGATGATG 1122
    |||||
Db 361 AlaArgValLeuAlaGluAlaMetSerGlnAlaThrAsnSerAlaAlaAlaValMetMet 380
OY 1123 CAGAAAGCAACTTCAAGGGCGCCCGCGCATCGTCAAGTGTTCAACTGCGGCAAGAG 1182
    |||||
Db 381 GlnArgGlyAsnPheLysGlyProArgLysThrIleLysCysPheAsnCysGlyLysGlu 400
OY 1183 GGCACATCGCGCGCAACTGCGCGCGCGCGCGCGCGAGAGAGGGCGTGTGAAGTGGCGAAG 1242
    |||||
Db 401 GlyHisIleAlaLysAsnCysValArgAlaProArgArgLysGlyCysTrpLysCysGlyLys 420
```

```
OY 1243 GAGGCCACCAAGATGAAGACTGACCGAGCGGCCCAACTTCTGGGCAAGATCTGG 1302
    |||||
Db 421 GluGlyHisGlnLeuLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrp 440
OY 1303 CCCAGCCACAAGGGCGCCCGCGCAACTTCTGAGAGCGCGCGGAGCCCGACCGCCCC 1362
    |||||
Db 441 ProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPro 460
OY 1363 CCCCGCGAGAGCTTCCGCTTC---GAGGAGACCAACCCCGCGCGCAAGAGAGCAAG 1419
    |||||
Db 461 ProAlaGlnSerPheGlyPheGlyGluGluIleThrProSerGlnLysGlnGluLys 480
OY 1420 GACCGCGAGACCTG-----ACCAGCCTGAAGAGCCTGTTGCGCAACGACCCCTGAGC 1473
    |||||
Db 481 AspLysGluLeuTyrProSerThrAlaLeuLysSerLeuPheGlyAsnAspProLeuLeu 500
OY 1474 CAG 1476
    |||
Db 501 Gln 501
```

## RESULT 2

FOLJND  
gag polyprotein - human immunodeficiency virus type 1 (Isolate NDK)  
N/Alternate names: core polyprotein  
N/Contains: core protein p15; core protein p17; core protein p24  
C/Species: human immunodeficiency virus type 1, HIV-1  
A/Note: host Homo sapiens (man)  
C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C/Accession: JQ0065  
R/Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.,  
Gene 81, 275-284, 1989  
A/Title: Nucleotide sequence of HIV-1-NDK: a highly cytopathic strain of the human immuno-  
A/Reference number: JQ0065; MUID:90034200; PMID:2806917  
A/Accession: JQ0065  
A/Molecule type: DNA  
A/Residues: 1-497 <SPI>  
A/Cross-references: UNIPROT:P18800; GB:M27323; NID:g328154; PIDN:AAA44868.1; PID:g328157  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polyprotein  
C/Keywords: AIDS; core protein; immunodeficiency; polyprotein  
F,1-129/Product: core protein p17 #status predicted <C17>  
F,130-389/Product: core protein p24 #status predicted <C24>  
F,390-497/Product: core protein p15 #status predicted <C15>

Alignment Scores:  
Pred. No.: 2,08e-109 Length: 497  
Score: 2294.50 Matches: 428  
Percent Similarity: 93.57% Conservative: 38  
Best Local Similarity: 85.94% Mismatches: 25  
Query Match: 82.77% Indels: 7  
DB: 1 Gaps: 4

US-09-475-704A-3 (1-1479) x FOLJND (1-497)

```
OY 1 ATGGGGCGCGCGCGCGCATCTGCGCGGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60
    |||||
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLysLeuAspThrTrpGluArgIleArg 20
OY 61 CTGCGCGCGCGCGCGCAAGAGTGTACTACATGATGAAGCACCTGTGTGGCGCAGCGCGAG 120
    |||||
Db 21 LeuArgProGlyGlyLysLysLysTyrAlaLeuLysHisLeuIleTrpAlaSerArgGlu 40
OY 121 CTGAGAGAGTTGCGCCTGAACCCCGCGCTGCTGAGAGACCGAGCGAGAGCTGAAGACCTGTTCAAC 180
    |||||
Db 41 LeuGluArgPheThrLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
OY 181 ATCCGCCAGCTGCACCCCGCGCTGACAGCCGCGAGAGAGCTGAAGACCTGTTCAAC 240
    |||||
Db 61 ILeGlyInLeuGlnProSerIleGlnThrGlySerGlnGluIleArgSerLeuTyrAsn 80
OY 241 ACCGTGGCCACCTGTACTGCGGTGCACGAGAGATCGAGTCCGCGACCAAGAGGCC 300
    |||||
```

D	b	81	ThrValAlaThrLeuTyrCysValHisGluArgIleGluValLysAspThrLysGluAla	100
Q	y	301	CTGGACAAGATCGAGGAGGAGAGACAAGTGGCCAGAGAGATCCAGAGGCCGAGGCC	360
			::::::::::	
D	b	101	ValGluLysMetGluGluGluGlnAsnLysSerLysLysThrGlnGlnAla--Ala	119
Q	y	361	GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGCAAGGCCAGATG	420
			::::::::::	
D	b	120	AlaAspSerSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet	139
Q	y	421	GTGGACACGAGGCCATCAGCCCCCGGACCCCTGAACGCGCTGGTGGAAGGTGATCGAGGAAG	480
D	b	140	ValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValIleGluGluLys	159
Q	y	481	GCCTTCAGCCCCCGAGGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCCACCCCCAG	540
D	b	160	AlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGluAlaThrProGln	179
Q	y	541	GACCTGAACACGATGTTGAACACCGTGGCGGCGCCACGAGCGCCATGACAGTGTGAAG	600
D	b	180	AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys	199
Q	y	601	GACACCATCAACGAGGAGGCGCGCGAGTGGAGCCGCGGTGACCCCGTGACGCGCGGCC	660
D	b	200	GluThrIleAsnAspGluAlaGluTyrPaspArgLeuHisProValHisGlyPro	219
Q	y	661	ATCGCCCCCGGCGAGATGCGCGGAGCCCCGCGGCGAGCAGACATCGCGGCAACGACACC	720
D	b	220	ValAlaProGluGlnMetArgGluProArgLysSerAspIleAlaGlyThrThrSerThr	239
Q	y	721	CTGCAGAGCAGATCGCCTGTGATGACCAGCAACCCCCCATCCCGTGGCGCATTTAC	780
D	b	240	LeuGlnGluGlnIleAlaTyrMetThrSerAsnProIleProValGlyGluIleTyr	259
Q	y	781	AAGCGGTGATCATCTCGGCGCTGAAACAAGATCGTCGGATGTAACGCCCGTGAGCATC	840
D	b	260	LysArgTyrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle	279
Q	y	841	CTGGACATCAAGCAGGCGGCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTCAAG	900
D	b	280	LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPheTyrLys	299
Q	y	901	ACCCTGCGCGCGAGCAGAGCACCCAGAGGTGAAGAACTGGATGACCGACACCTGCTG	960
D	b	300	ThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsnTyrMetThrGluThrLeuLeu	319
Q	y	961	GTGCAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCTTCGGCCCCGCGCGCAGC	1020
D	b	320	ValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGlnAlaThr	339
Q	y	1021	CTGGAGGAGATGATGACCGCCTGCGCAGGGCGTGGGGCGGCCACGACCAAGGCCGCGTG	1080
D	b	340	LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGluHisLysAlaArgVal	359
Q	y	1081	CTGGCCGAGGCGGATGAGCCAGGCCAAC-----ACCAGCGTGATGATGACAAGAGC	1131
D	b	360	LeuAlaGluAlaMetSerGlnValThrGlySerAlaThrAlaValMetMetGlnArgGly	379
Q	y	1132	AACTTCAGGGGCCCCCGCGCATCGTCAAGTGTCTCAACTGGCGCAAGAGGCCACATC	1191
D	b	380	AsnPheLysGlyProArgLysSerIleLysCysPheAsnCysGlyLysGluGlnHisThr	399
Q	y	1192	GCCCGCAACTGCGCGCGCCCCCGGCAAGAGGCGCTGTGAAGTGGCGGCAAGAGGGCCAC	1251
D	b	400	AlaLysAsnCysArgAlaProArgLysLysGlyCysTyrLysCysGlyArgGluGlnHis	419
Q	y	1252	CAGATGAAGGACTGCAACGAGCGCCAGGCCAACTTCTGGGCAAGATTTGGCCAGCCAC	1311
D	b	420	GlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTyrProSerHis	439
Q	y	1312	AAGGGCGCGCGCGCAACTTCTGTGACAGCGCGCGCGCGGACCCAGCCCGCGCGGAG	1371
D	b	440	LysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGlu	459

```

QY      1372 AGCTTCGGCTTC---GAGGAGACCACCCCGGCCAGAAAGCAGAGAACGACCGGCAG 1428
        |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      460 SerPheGlyPheGlyGluGluIleThrProSerGlnGlnGlnGlnGlnValAspLysGln 479
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      1429 -----ACCTTGACCAGCCTGTGAAGAGCCTGTTCCGCAACGACCCCTTGAGCCAG 1476
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      480 LeuTyrProLeuAlaSerLeuLysSerLeuPheGlyAsnAspProSerSerGln 497
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 3
T01667
gag polyprotein - human immunodeficiency virus type 1
C/Species: human immunodeficiency virus type 1, HIV-1
C/Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 09-Jul-2004
C/Accession: T01667
R/Alixon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.
Cell 46, 63-74, 1986
A/Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol
A/Reference number: Z14389; MUID:86245056; PMID:2424612
A/Accession: T01667
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: mRNA
A/Residues: 1-505 <Alt>
A/Cross-references: UNIPROT:P04594; EMBL:K03456; NID:g60228; PIDN:CAA28011.1; PID:g60223
C/Superfamily: AIDS-related virus gag polyprotein

Alignment Scores:
Pred. No.:          5.51e-108           Length:          505
Score:              2266.50             Matches:           431
Percent Similarity: 90.71%               Conservative:      28
Best Local Similarity: 85.18%            Mismatches:       32
Query Match:        81.76%               Indels:           15
DB:                  2                   Gaps:             5

US-09-475-704A-3 (1-1479) x T01667 (1-505)

QY      1 ATGGGCGCGCCGCGCACGATCCTGCGCGCGCGCAAGCTGAGCGCCTTGAGCGCATCCGC 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLysLeuAspAlaTrpGluLysIleArg 20
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      61 CTGCGCCCCGCGCGCAAGAAGTGCTACATGATGAAGCACCTGTGTGGCCAGCCGCGAG 120
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      21 LeuArgProGlyGlyLysLysLysTyArgLeuLysHisLeuValTrpAlaSerArgGlu 40
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      121 CTGAGAAAGTTGCGCCTGTAACCCCGGCGCTGTGAGACCGAGGCGCTGCAAGCAGATC 180
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluIleThrGlyGluGlyCysGlnGlnIle 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      181 ATCCGCCAGCTGCACCCCGCCCTGCAGACCGCGCAGCGAGACTGAAGACCTGTTCAAC 240
        ::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      61 MetGluGlnLeuGlnSerThrLeuLysTrnGlySerGluGluIleLysSerLeuTyrAsn 80
        ::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      241 ACCGTGGCCACCTGTACTGCGGTGCACGAGAAAGATCGAGSTCCGCGACCAAGAGGCC 300
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      81 ThrValAlaThrLeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAla 100
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      301 CTGACAAGATCGAGAGGACGACAAGAACTGCCAGCAAGATCCAGCAGGCGCGAGGCC 360
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      101 LeuAspLysIleGluGluIleGlnAsnLysSerArgGlnLysThrGlnGlnAlaAlaAla 120
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      361 GCC-----GACAAAGGCGAAGGTGAGCCAGCACTACCCCATC 396
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121 AlaGlnGlnAlaAlaAlaAlaThrLysAsnSerSerValSerGlnAsnTyrProIle 140
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      397 GTGCAGAACCTGCAGGGCGCAGATGTTGCACCGCCATCAGCCCCCGCACCTGAACGCC 456
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      141 ValGlnAsnAlaGlnGlyGlnMetIleHisGlnAlaIleSerProArgThrLeuAsnAla 160
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      457 TGGGTGAAGGTGATCGAGAGAAAGCGCTTCAGCCCCGAGGTGATCCCATGTTACCGGCC 516
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      161 TrpValLysValIleGluGluLysAlaPheSerProGluValIlePrometPheSerAla 180
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      517 CTGAGCGAGGGCGCCACCCCGCAGACCTGAACACGATGTTGAACACCGTGGCGGCCAC 576

```

```
|||||
Db      181 LeuSerGluGlyAlaThrProGlnAspLeuAsnMetMetLeuAsnIleValGlyHis 200
QY      577 CAGCGCCCATGACAGTGTGAAGGACACCATCAAGAGAGGCGCCGAGTGGAACCGC 636
Db      201 GlnAlaIleMetGlnMetLeuLysAspThrIleAsnGluGlnAlaIleAspTyrAspArg 220
QY      637 GTGACACCCCGTGACAGCCGCGCCCATCGCCCGGAGATGCGCGGAGCCCGCGGAGC 696
Db      221 ValHisProValHisAlaGlyProIleProGlyGlnMetArgGluProArgGlySer 240
QY      697 GACATCGCCGCGACCAACAGACCCCTGACAGAGACGATCGCCTGATGACCAACCCC 756
Db      241 AspIleAlaGlyThrThrSerThrLeuGlnGlnIleGlyTyrMetThrSerAsnPro 260
QY      757 CCATCCCGCGTGAGCATCTACAAGCGGTGATCATCTGGCGCTGAACAAGATCGTG 816
Db      261 ProIleProValGlyAspIleTyrLysArgTyrIleIleLeuGlyLeuAsnLysIleVal 280
QY      817 CGGATGTACAGCCCGGTGAGCATCTCGGACATCAAGCAGGCGCCCAAGAGCCCTTCCGC 876
Db      281 ArgMetTyrSerProValSerIleLeuAspIleArgGlnGlyProLysGluProPheArg 300
QY      877 GACTACGTGAGCCGCTTCTCAAGACCCCTGCGCGCGGACGACAGCAACCCAGAGGTGAAG 936
Db      301 AspTyrValAspArgPhePheLysThrLeuArgAlaGluGlnAlaThrGlnGlnValLys 320
QY      937 AACTGGATGACCGACACCTGCTGTGGTGACAGACCGCAACCCCGACTGCAAGACCATCTG 996
Db      321 AsnTyrMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeu 340
QY      997 CGCGCTCTCGCGCGCGCGGCGGACGCTGAGAGAGATGATGACCGCCTGCCAGGCGTGCGC 1056
Db      341 LysAlaIleGluGlyProGlyAlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGly 360
QY      1057 GGCCCCAGCCACAGGCGCGCGCTGCTGGCGGAGCGCATGAGCCAGCCACACCAACG--- 1113
Db      361 GlyProSerHisLysValAlaArgValLeuAlaGlnAlaMetSerGlnAlaThrAsnSerThr 380
QY      1114 -----GTGATGATGACAGAGGCAACTTCAAGGCGCGCGCATCGTCAAGTCTTC 1167
Db      381 AlaAlaIleMetMetGlnArgGlyAsnPheLysGlyGlnLysArgIle---LysCysPhe 399
QY      1168 AACTGCGGCAAGAGAGGCGCACATCGCCCGGCACTGCGCGCGCGCGCGCAAGAGGCTGC 1227
Db      400 AsnCysGlyLysGlnGlyHisIleValAlaArgAsnCysArgAlaProArgLysLysGlyCys 419
QY      1228 TGGAAGTGCGGCAAGAGGCGCACAGATGAAGACTGCAACCGGCGCGCAACTTC 1287
Db      420 TrrLysCysGlyLysGlnGlyHisIleMetLysAspCysThrGlnArgGlnAlaAsnPhe 439
QY      1288 CTGGGCAAGATCTGGCCCAAGCCACAAAGGCGCGCGCGCAACTTCTGACAGAGCGCGCCC 1347
Db      440 LeuGlyLysIleTrrProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgPro 459
QY      1348 GAGGCCACCGCGCGCGCGCGCGGAGAGCTTCCGCTTC--GAGGAGACCAACCCCGGCGCAG 1404
Db      460 GluProThrAlaProProAlaGlnSerPheGlyGlnIleLysProSerGln 479
QY      1405 AAGCAGAGAGAGAGAGACCGCGAG-----ACCTTGACCAAGCCTGAAGAGCCTGTTCGGC 1458
Db      480 LysGlnGlnGlnLysAspLysGlnLeuTyrProLeuAlaSerLeuLysSerLeuPheGly 499
QY      1459 AAGCAGACCCCTGAGCCAG 1476
Db      500 AsnAspGlnLeuSerGln 505
```

```
RESULT 4
FOVWLV
gag polyprotein - human immunodeficiency virus type 1 (isolate LAV-1a)
N/Alternate names: assemblin; core polyprotein; gag precursor
N/Contains: capsid antigen core protein p24CA; core protein p1; core protein p6; matrix
C/Species: human immunodeficiency virus type 1, HIV-1
```

```
A/Note: host Homo sapiens (man)
C/Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 05-Dec-1998
C/Accession: A03946
R/Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A/Title: Nucleotide sequence of the AIDS virus, LAV.
A/Reference number: A90866; MUID:8509333; PMID:2981635
A/Accession: A03946
A/Molecule type: DNA
A/Residues: 1-500 <MAT>
C/Genetics:
A/Gene: gag
C/Superfamily: AIDS-related virus gag polyprotein
C/Keywords: AIDS; blocked amino end; core protein; immunodeficiency; lipoprotein; myrist
F/2-500/Product: gag precursor (assemblin) #status predicted <GAG>
F/2-132/Product: matrix antigen core protein p17MA #status predicted <P17>
F/20-32/Region: nuclear location signal
F/110-114/Region: nuclear location signal
F/133-363/Product: capsid antigen core protein p24CA #status predicted <P24>
F/364-377/Product: core protein p2 #status predicted <CP2>
F/378-432/Product: nucleocapsid core protein p7NC #status predicted <CP7>
F/392-405/Region: zinc finger CCHC motif
F/413-426/Region: zinc finger CCHC motif
F/433-448/Product: core protein p1 #status predicted <CP1>
F/449-500/Product: core protein p6 #status predicted <CP6>
F/2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted
F/392,395,400,405/Binding site: zinc (Cys, Cys, His, Cys) #status experimental
F/413,416,421,426/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

Alignment Scores:
Pred. No.:      8.29e-108      Length:      500
Score:          2263.00      Matches:      425
Percent Similarity: 92.61%      Conservative: 39
Best Local Similarity: 84.83%      Mismatches:  27
Query Match:    81.64%      Indels:      10
DB:             1          Gaps:          5

US-09-475-704A-3 (1-1479) x FOVWLV (1-500)
QY      1  ATGGGCGCGCGCGGCGGACATCTCGCGCGCGCAAGCTGACGCGCTGGAGCGCATCCGC 60
Db      1  MetGlyAlaIleArgAlaSerValLeuSerGlyGlyGlnLeuAspArgTrrGlyLysIleArg 20
QY      61  CTGCGCGCGCGCGGCAAGAGTGCTACATGATGAAGCACTGTGTGGCCAGCGCGAG 120
Db      21  LeuArgProGlyGlyLysLysLysTrrLysLeuLysHisIleValTrrAlaSerArgGln 40
QY      121  CTGGAAGAAGTTCGCCCTGAACCCCGGCGCTGTGAGAGACCAAGGCGGTGCAAGCATC 180
Db      41  LeuGlnArgPheAlaValaAsnProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60
QY      181  ATCCGCCAGCTGCACCCCGCGCTGCAGACCGGCGAGGAGCTGAAGAGCCTGTCAAC 240
Db      61  LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnLeuArgSerLeuTyrAsn 80
QY      241  ACCGTGGCCACCTGTACTGGGTGCACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300
Db      81  ThrValAlaThrLeuTyrCysValHisGlnArgIleGlnIleLysAspThrLysGlnAla 100
QY      301  CTGGAACAAGATCGAGAGGAGCAGAAACAAGTGCACGACAGAAAGATCCAGAGGCCGAGGCC 360
Db      101  LeuAspLysIleGlnGlnGlnAsnLysSerLysLysLysAlaGlnGlnAla--Ala 119
QY      361  GCCGACAAGGGC-----AAGGTGAGCGCAAGTACCCTCATCGTGAGAACTGCAG 411
Db      120  AlaAspThrGlyHisSerSerGlnValSerGlnAsnTrrProIleValGlnAsnIleGln 139
QY      412  GGCAGAGTGTGCACCAAGGCGCATCAGCCCGCGACCCCTGAAAGCGCTGGTGAAGGTGATC 471
Db      140  GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrrValLysValVal 159
QY      472  GAGGAGAAGGCGCTTCAAGCCCGAGGTGATCCCATGTTCAACGCGCCTGAGCGAGGCGCC 531
```

Db 160 GluGluValAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179  
QY 532 ACCCCCGAGAACTGAACAGATGTTGAACACCGTGGCGGCCACAGCGCCCATGCGAG 591  
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaMetGln 199  
QY 592 ATGCTGAAGACACCATCAACGAGAGCGCGCCGAGTGGGACCGCGTGACCCCGTGAC 651  
Db 200 MetLeuValSerGluThrIleAsnGluGlnAlaGluValTrpAspArgValHisProValHis 219  
QY 652 GCCGGCCCCATCGCCCCCGCCAGATCGCGGAGCCCGCGGACGACATCGCGGAC 711  
Db 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239  
QY 712 ACCAGCACCTGACAGAGAGATCGCTGGATGACAGAACCCCGCATCCCGTGCGG 771  
Db 240 ThrSerThrLeuGlnGlnIleGlyTrpMetThrAsnAsnProProIleProValGly 259  
QY 772 GACATCTAACAGCGGTGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCC 831  
Db 260 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279  
QY 832 GTGAGCATCTGACATCAAGCAGGCGCCCAAGAGCGCTTCGCGACTACGTGACCG 891  
Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299  
QY 892 TTCTTCAAGACCTCGCGCGCCGAGCAGACCAACCAGAGGTGAAGACTGATGACCG 951  
Db 300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnIleValLysAsnTrpMetThrGlu 319  
QY 952 ACCCTGCTGTGAGAAAGCCAAACCCCGACTGCAAGACATCTCGCGCTCTCGGCC 1011  
Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339  
QY 1012 GCGCGACCTGAGAGATGATGACCGCTGCGAGGCGGTGGCGGCCAGCCACAG 1071  
Db 340 AlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359  
QY 1072 GCCCGGTGCTGGCGGAGCGATGAGCCAGCCAAACACAGC-----GTGATGATGAG 1125  
Db 360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGln 379  
QY 1126 AAGAGCACTTCAAGGGCCCCGCGCATGCTCAAGTCTCAACTGCGCAAGAGGCG 1185  
Db 380 ArgGlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnGlyLysGluGly 399  
QY 1186 CACATGCGCCGCAACTGCGCGCGCCCGCCGCAAGAGGCGTGTGAAGTGGCAAGAG 1245  
Db 400 HisIleAlaArgAsnGlyAlaProArgLysLysGlyCysTrpLysCysGlyLysGlu 419  
QY 1246 GGCACACGATGAAGGACTGACACCGAGCGCCAGGCAACTTCTGGGCAAGATCTGGCC 1305  
Db 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439  
QY 1306 AGCCACAAGGGCGCGCCGCAACTTCTGACAGAGCGCGCCGAGCCACCGCCCGCC 1365  
Db 440 SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459  
QY 1366 GCCGAGACTTCCGC-----TTCGAGAGAGACACCGCCGCGCAGAGCAGAGAG 1419  
Db 460 GluGluSerPheArgSerGlyValGluThrThrProSerGlnLysGlnIleProIle 479  
QY 1420 GACCGCGAG-----ACCTGACACGCTGAAGAGCCTGTCGCAACGACCCCGTGAGC 1473  
Db 480 AspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsnAspProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500

RESULT 5  
FOVWH4  
gag polyprotein - human immunodeficiency virus type 1 (isolate CDC-451)

N:Alternate names: core polyprotein  
N:Contains: core protein p15; core protein p17; core protein p24  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 09-Jul-2004  
C:Accession: A25523  
R:Desai, S.M.; Kalyanaraman, V.S.; Casey, J.M.; SriLivasan, A.; Andersen, P.R.; Devare, Proc. Natl. Acad. Sci. U.S.A. 83, 8380-8384, 1986  
A:Title: Molecular cloning and primary nucleotide sequence analysis of a distinct human A:Reference number: A94136; MUID:87041461; PMID:3490666  
A:Accession: A25523  
A:Molecule type: DNA  
A:Residues: 1-500 <DES>  
A:Cross-references: UNIPROT:P05887; GB:M13136; NID:g326459; PIDN:AAA44306.1; PID:g32646.  
C:Genetics:  
A:Gene: gag  
C:Superfamily: AIDS-related virus gag polyprotein  
C:Keywords: core protein; polyprotein  
F:1-132/Product: core protein p17 #status predicted <P17>  
F:133-391/Product: core protein p24 #status predicted <P24>  
F:392-500/Product: core protein p15 #status predicted <P15>

Alignment Scores:  
Pred. No.: 4.27e-107 Length: 500  
Score: 2249.00 Matches: 422  
Percent Similarity: 92.22% Conservative: 40  
Best Local Similarity: 84.23% Mismatches: 29  
Query Match: 81.13% Indels: 10  
DB: 1 Gaps: 5

US-09-475-704A-3 (1-1479) x FOVWH4 (1-500)

QY 1 ATGGGCGCGCGCGCCAGCATCTGCGCGCGCGCAAGCTGACCGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTrpGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGCAAGAGTGTCTATGATGAGCACTGTGTGGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysGlnTyrArgLeuLysHisIleValTrpAlaSerArgLys 40  
QY 121 CTGGAAGATTCGCCCTGAACCCCGCGCTGTGAGACCAAGCGGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaValAlaAsnProGlyLeuLeuGluThrSerLysGlyCysArgGlnIle 60  
QY 181 ATCCGCCAGCTGCACCCCGCCCTGACAGCCGCGCAGGAGGAGTGAAGCCTGTTCAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGAACGAGAGATCGAGGTCCCGGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGCAGAAACAAGTCCAGCAGAGATTCAGAGGCCGAGGCC 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysAlaGlnGlnAla---Ala 119  
QY 361 GCCGACAAGGGC-----AAGGTGAGCCAGAACTACCCCATCTGTGACAGACTGCGAG 411  
Db 120 AlaAspThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGln 139  
QY 412 GGCAGATGTGACACGAGGACATGACCCCGCACCCTGAACGCTGTGGTGAAGTGATC 471  
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIle 159  
QY 472 GAGGAGAAGGCTTTCAGCCCGGAGGTATCCCATGTTCACCGCGCTGAGCGAGGGGCC 531  
Db 160 GluGluLysAlaPheSerProGluValIleProMetPheAlaAlaLeuSerGluGlyAla 179  
QY 532 ACCCCCGAGACCTGAACACGATGTTGAACACCGTGGCGGCGCAGCGCGCATGCGAG 591  
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 199  
QY 592 ATGCTGAAGACACCATCAACGAGAGCGCGCGAGTGGGAGCCGCTGACCCCGTGAC 651

Db 200 MetLeuLysGIuThrIleAsnGIuGluAlaIleGIuTrpAspArgLeuHisProValHis 219  
QY 652 GCCGGCCCCATCGCCCCGGCCAGATGGCGGAGCCCCGGCGGACGACATCGCCGGCACC 711  
Db 220 AlaGIYProIleAlaProGIuGlnMetArgGIuProArgGlySerAspIleAlaGIYThr 239  
QY 712 ACCAGCACCTTGACAGAGACAGATCGCCTGGATGACCAACACCCCCCATCCCGTGGGC 771  
Db 240 ThrSerThrLeuGlnGlnGlnIleGIYTrpMetThrAsnProProThrProValGIY 259  
QY 772 GACATCTAACAGCGGTGATCATCTTGCGCTGAACAAGATCGTGGGATGTACAGCCCC 831  
Db 260 GIuIleTYrLysArgTrpIleIleLeuGIYLeuAsnLysIleValArgMetTYrSerPro 279  
QY 832 GTGAGCATCTTGACATCAAGCAGGGGCCCAAGAGAGCCCTTCGGCGACTACGTGACCGC 891  
Db 280 IleSerIleLeuAspIleArgGlnGIYProLysGIuProPheArgAspTYrValAspArg 299  
QY 892 TTCTTCAAGACCCCTGCGCGCCGAGCAGAGACACCCAGAGGTGAAGAACTGGATGACCGAC 951  
Db 300 PheTYrLysThrLeuArgAlaGIuGlnAlaSerGlnGIuValLysAsnTrpMetThrGlu 319  
QY 952 ACCCTGCTGTGACAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGCCCC 1011  
Db 320 ThrLeuLeuValGIuAsnAlaAsnProAspCysLysThrIleLeuValAlaLeuGIYPro 339  
QY 1012 GCGCGCACCTTGAGAGATGATGACCGCCTGCGAGGCGGTGGCGGCCAGCCACAAG 1071  
Db 340 AlaAlaThrLeuGIuGlnMetMetThrAlaCysGlnGIYValGIYGIYProGIYAsnLys 359  
QY 1072 GCCCGCGTGTGGCCGAGCGATGAGCCAGGCCAACACCAGC-----GTGATGATGCAG 1125  
Db 360 AlaArgValLeuAlaGIuAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGln 379  
QY 1126 AAGAGCAACTTCAAGGGCCCCCGCGCATCGTCAAGTGCTTCACTGGCGGCAAGAGGGGC 1185  
Db 380 ArgGIYAsnPheArgArgGlnGIYLysThrValLysCysPheAsnCysGIYLysGIuGIY 399  
QY 1186 CACATGCCCCGCAACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGGAG 1245  
Db 400 HisIleAlaArgAsnCysLysAlaProArgLysGIYLysCysTrpLysCysGIYArgGIu 419  
QY 1246 GGGCCACCATGAAGACTGCACCGAGCGGCCCAAGGCACTTCTGGGCAAGATCTGGCCC 1305  
Db 420 GIYHisGlnMetLysAspCysThrGIuArgGlnAlaAsnPheLeuGIYLysIleTrpPro 439  
QY 1306 AGCCACAAGGGCGCGCGCGCAACTTCTGCAGAGCGCGCGCGCCAGGCCACCGCCCCCCC 1365  
Db 440 SerHisLysGIYArgProGIYAsnPheLeuGlnSerArgProGIuProThrAlaProPro 459  
QY 1366 GCCGAGAGCTTCCGCTTC-----GAGGAGACACACCCCGGCCAGAGCAGAGAGCAAG 1419  
Db 460 GIuGIuSerPheArgPheGIYAspGIuThrThrProSerGlnLysGIuGlnProArg 479  
QY 1420 GACCGCGAG-----ACCCTGACCAAGCTGAAGAGCCTGTTCGGCAACGACCCCTGAGC 1473  
Db 480 AspLysGIuLeuTYrProLeuAlaSerLeuArgSerLeuPheGIYAsnAspProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500

RESULT 6  
FOVMA2  
gag polyprotein - human immunodeficiency virus type 1 (isolate ARV-2)  
N/Alternate names: core polyprotein  
N/Contains: core protein p15; core protein p17; core protein p24  
C/Species: human immunodeficiency virus type 1, HIV-1  
A/Note: host Homo sapiens (man)  
C/Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004  
C/Accession: A03947  
R/Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-SH  
Science 227, 484-492, 1985

A/Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).  
A/Reference number: A04003; MUID:85090453; PMID:2578227  
A/Accession: A03947  
A/Molecule type: DNA  
A/Residues: 1-502 <SAN>  
A/Cross-references: UNIPROT:P03349; GB:K02007; NID:g328658; PIDN:AAB59875.1; PID:g328661  
C/Comment: Cleavage sites that yield the mature core proteins remain to be determined.  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polyprotein  
C/Keywords: AIDS; core protein; immunodeficiency; polyprotein  
F/1-134/Product: core protein p17 #status predicted <p17>  
F/135-393/Product: core protein p24 #status predicted <p24>  
F/394-502/Product: core protein p15 #status predicted <p15>

Alignment Scores:  
Pred. No.: 5,39e-107 Length: 502  
Score: 2247.00 Matches: 418  
Percent Similarity: 92.43% Conservative: 46  
Best Local Similarity: 83.27% Mismatches: 28  
Query Match: 81.06% Indels: 10  
DB: 1 Gaps: 4

US-09-475-704A-3 (1-1479) x FOVMA2 (1-502)

QY 1 ATGGGCGCGCGCCAGCATCTGCGCGCGGCGCAAGCTGAGCGCCTGGAGCCCATCCGC 60  
Db 1 MetGIYAlaArgAlaSerValLeuSerGIYGIYLeuAspLysTrpGIYLysIleArg 20  
QY 61 CTGGCGCGCGCGCGCAAGAAGTGTACATGATGAAGCACCTGTGTGGCGCCCGCGAG 120  
Db 21 LeuArgProGIYGIYLysLysLysTYrLysLeuLysHisIleValTrpAlaSerArgGIu 40  
QY 121 CTGGAAGAGTTCGCCCTGAACCCCGCGCTGTGAGAGACCAAGAGGCGTGAAGCATC 180  
Db 41 LeuGIuArgPheAlaValAlaAsnProGIYLeuLeuGIuThrSerGIuGIYCysArgGlnIle 60  
QY 181 ATCCGCCAGCTGCACCCCGCGCTGCAGACCGGACGAGAGAGCTGAAGAGCCTGTTCAAC 240  
Db 61 LeuGIYLysLeuGlnProSerLeuGlnThrGIYSerGIuLeuLysArgSerLeuTYrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGACAGAGAGATCGAGGTCCGACACCAAGAGAGGCC 300  
Db 81 ThrValAlaThrLeuTYrCysValHisGlnArgIleAspValLysAspThrLysGIuAla 100  
QY 301 CTGGACAAGATTCGAGAGAGAGAGACAAAGTCCAGCAGAGATCCAGAGCCCGAGGCC 360  
Db 101 LeuGIuLysIleGIuGIuGlnAsnLysSerLysLysAlaGlnGlnAlaAlaAla 120  
QY 361 GCC-----GACAAGGGCAAGGTGAGCCAGACTACCCCATCTGTGCAGAACCTG 408  
Db 121 AlaAlaGIYThrGIYAsnSerSerGlnValSerGlnAsnTYrProIleValGlnAsnLeu 140  
QY 409 CAGGGCCAGATGTGTGACACCGCCATCAGCCCGCCGACCTGAACGCGCTGGAAGGTG 468  
Db 141 GlnGIYLysMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysVal 160  
QY 469 ATCGAGGAGAGAGCCTTCAGCCCGCGAGGTATCCCATGTTACCGCCCTGAGCGAGGGC 528  
Db 161 ValGIuGIuLysAlaPheSerProGIuValIleProMetPheSerAlaLeuSerGIuGIY 180  
QY 529 GCCACCCCGCCAGACCTGAACAAGATGTTGAACACCGTGGCGGCCACCAAGCCCGCATG 588  
Db 181 AlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGIYHisGlnAlaAlaMet 200  
QY 589 CAGATGCTGAAGACACCATCAACGAGAGCGCCCGAGTGGGACCGCGTGCACCCCGTGC 648  
Db 201 GlnMetLeuLysGIuThrIleAsnGIuGlnAlaAlaGIuTrpAspArgValHisProVal 220  
QY 649 CACGCGCGCCCATCGCCCCCGCGCAGATGCGCGAGCCCGCGCGCAGCAGCATCGCCGCGC 708  
Db 221 HisAlaGIYProIleAlaProGIYLysMetArgGIuProArgGIYSerAspIleAlaGIY 240

```
QY 709 ACCACGACGACCTGACAGACAGATCGCTTGATGACCAACCCCGCATCCCGCTG 768
Db 241 ThrThrSerThrLeuGlnGlnIleGlyTyrMetThrAsnAsnProIleProVal 260
QY 769 GGGACATCTCAAGCGGTGGATCATCTGGGGCTGAACAAGATCGTGGGATGTACAG 828
Db 261 GtGtIleTyrLysArgTyrIleIleLeuGtLysValAsnLysIleValArgMetTyrSer 280
QY 829 CCGGTGAGCATCTTGACATCAAGCAGGCGCCCAAGAGCGCTTCCGGCACTACGTGAC 888
Db 281 ProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAsp 300
QY 889 CGCTTCTCAAGACCTGCGCGCGCCGAGCAGACCCAGAGGTGAAGACTGTGATGAC 948
Db 301 ArgPheTyrLysThrLeuArgAlaGlnAlaSerGlnAspValLysAsnTyrMetThr 320
QY 949 GACACCTGCTGTGTCAGAACGCCAACCCGACTGCAGAACCATCTGCGCGCTTCCG 1008
Db 321 GluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGly 340
QY 1009 CCGCGCGCCAGCTTGAGAGATGATGACCGCTGCAAGGGCGTGGCGCCCGACGAC 1068
Db 341 ProAlaAlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHis 360
QY 1069 AAGGCGCGCTGTGGCGGAGCGATGAGCCAGGCCAAC-----ACCAGCGTATGATG 1122
Db 361 LysAlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnProAlaAsnIleMetMet 380
QY 1123 CAGAGAGCAACTTCAAGGGCCCCCGGCGCATCTGTAAGTCTTCAACTGCGCAAGAG 1182
Db 381 GlnArgGlyAsnPheArgAsnGlnArgGlySerThrValLysCysPheAsnCysGlyLysGln 400
QY 1183 GGCACATGCGCCGCACTGCGCGCGCGCGCGCGCAAGAGGGCTGTGAGTGGCGCAAG 1242
Db 401 GtHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTyrArgCysGlyArg 420
QY 1243 GAGGCGCACAGATGAAGACTGCACGAGCGCGCAACTTCTGGGCAAGATGTGG 1302
Db 421 GtGtLysHisGlnMetLysAspCysThrGtLysArgGlnAlaAsnPheLeuGtLysIleTyr 440
QY 1303 CCGACGACCAAGGGCGCGCGCGCGCAACTTCTGACAGCGCGCGCGCGCGCGCGCC 1362
Db 441 ProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPro 460
QY 1363 CCGCGCGAGAGCTTCGGCTTC-----GAGGAGACCAACCCCGCGCAAGACGAGAGC 1416
Db 461 ProGlnLysSerPheArgPheGtGtLysGtLysThrThrProSerGlnLysGlnGluPro 480
QY 1417 AAGACCGCGAG-----ACCCTGACGAGCTTGAAGAGCTTGTGGCAACGACCCCTG 1470
Db 481 IleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGtLysAsnAspProSer 500
QY 1471 AGCCAG 1476
Db 501 SerGln 502
```

## RESULT 7

```
FOVWH3
gag polyprotein - human immunodeficiency virus type 1 (isolate HTLV-III, BH10)
N:Alternate names: core polyprotein
C:Contains: core protein p15; core protein p17; core protein p24
C:Species: human immunodeficiency virus type 1, HIV-1
A>Note: host Homo sapiens (man)
C:Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C:Accession: A03945
R:Rather, L.; Haseltine, W.; Patarca, R.; Livak, K.J.; Starcich, B.; Josephs, S.F.; Doranberger, J.A.; Papas, T.S.; Ghayeb, J.; Chang, N.T.; Gallo, R.C.; Wong-Staal, F.
Nature 313, 277-284, 1985
A:Title: Complete nucleotide sequence of the AIDS virus, HTLV-III.
A:Reference number: A93353; MUID:8511123; PMID:2578615
A:Accession: A03945
A:Molecule type: DNA
A:Residues: 1-512 <RAT>
```

```
A:Cross-references: UNIPROT:P03347; GB:M15654; GB:K02008; GB:K02009; GB:K02010; NID:g32e
C:Comment: Cleavage sites that yield the mature core proteins remain to be determined.
C:Genetics:
A:Gene: gag
C:Superfamily: AIDS-related virus gag polyprotein
C:Keywords: AIDS; core protein; immunodeficiency; polyprotein
F:1-132/Product: core protein p17 #status predicted <P17>
F:133-391/Product: core protein p24 #status predicted <P24>
F:392-512/Product: core protein p15 #status predicted <P15>

Alignment Scores:
Pred. No.: 2.19e-106 Length: 512
Score: 2235.00 Matches: 424
Percent Similarity: 90.64% Conservative: 41
Best Local Similarity: 82.65% Mismatches: 26
Query Match: 80.63% Indels: 22
DB: 1 Gaps: 7

US-09-475-704A-3 (1-1479) x FOVWH3 (1-512)
QY 1 ATGGCGCGCGCGCGACATCTGCGCGCGCGCAAGCTGACGCGCTGGAGCGCATCGC 60
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLysLeuAspArgTyrGluLysIleArg 20
QY 61 CTGCGCGCGCGCGCAAGAAGTGTACATGATGAACACCTGTGTGGCGCAGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTyrAlaSerArgGlu 40
QY 121 CTGAGAAGTTGGCTTGAACCCCGCGCTGTGAGACCGAGCGCTGCAAGCAGATC 180
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGlnGlyCysArgGlnIle 60
QY 181 ATCGCGCGAGTGAACCCCGCGCTGACACCGCGAGCGAGAGCTGAAGACTGTTCAC 240
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGluLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCACCCTGTACTGTGTCGACGAGAAGATCGAGTCCGCGCACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100
QY 301 CTGCAAGATCGAGAGGAGGACAGAACAAAGTGCCAGCAGAAGATCCAGAGCGCGAGCC 360
Db 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 119
QY 361 GCCGACAGGGC-----AAGGTGAGCCAGAACTAACCCCATCTGACAGAACTGCAG 411
Db 120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139
QY 412 GGCAGATGTGACACGAGGCATACCCCGCGCACTGAACGCTGGGTGAAGTATC 471
Db 140 GtGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValVal 159
QY 472 GAGAGGAAGGCTTACGCGCGCGAGGTATCCCATGTTACCGCGCTGAGCGAGGCC 531
Db 160 GtGtLysValaPheSerProGluValIleProMetPheSerAlaLeuSerGlnGlyAla 179
QY 532 ACCCGCCAGGACCTGAACAGATGTTGAACACCGTGGCGCGCGCACGCGCCATGCAG 591
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGln 199
QY 592 ATGCTGAAGACACCATCAACGAGAGGCGCGCGAGTGGACCGCGTGCACCCCGTGCAC 651
Db 200 MetLeuLysGluThrIleAsnGlnGluAlaAlaGluTyrAspArgValHisProValHis 219
QY 652 GCCGCGCCCATCGCGCGCGCGCGAGATGCGGAGCGCGCGCGCGAGGACATGCGCGCAC 711
Db 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239
QY 712 ACCAGCACCTGCAAGAGAGCAGATGCTGTGATGACCAAGACCCCGCATCCCGTGGGC 771
Db 240 ThrSerThrLeuGlnGlnGlnIleGlyTyrMetThrAsnAsnProIleProValGly 259
QY 772 GACATCTCAAGCGGTGATCATCTGCGCTGGAACAAGATCGTGGGATGTACAGCCCC 831
```

Db 260 G|u|l|e|T|y|r|l|y|s|a|r|g|t|p|i|l|e|i|l|e|u|e|u|l|y|l|e|u|a|s|n|l|y|s|i|l|e|v|a|l|a|g|m|e|t|t|y|r|s|e|r|p|ro 279  
QY 832 GTGAGCATCCTTGACATCAAGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGACCGC 891  
Db 280 ThrSerIleLeuAspIleArgInglYProlySgluProPheArgAspTyrValAspArg 299  
QY 892 TTCTTCAGAACCTTGGCGCCCGAGCAGACACCAGAGGTGAAGACTGTGATGACCGAC 951  
Db 300 PheTyrIysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGlu 319  
QY 952 ACCCTGCTGTGACAGACGCCAACCCGACTGCAAGACCATCTGCGGCTCTCGGCC 1011  
Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysIysThrIleLeuYsaIalaLeuGlyPro 339  
QY 1012 GGCGCCAGCTTGAGAGATGATGACCGCTGCGAGGCGTGGCGGCCCAAG 1071  
Db 340 AlaAlaThrLeuGluGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359  
QY 1072 GCCCGGTGCTGGCCGAGCGATGAGCCAG--GCCAACACC--AGCGTATGATGCAG 1125  
Db 360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln 379  
QY 1126 AAGAGCACTTCAAGGCGCCCGCGCAATCGTCAAGTCTTCACTGCGGCAAGAGGCG 1185  
Db 380 ArgGlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnCySgIlySgIly 399  
QY 1186 CACATCGCCCGCAACTGCGCGCGCGCGCGCGCAAGAGGCTGCTGGAAGTGGCGAAGAG 1245  
Db 400 HisThrAlaArgAsnCySgIlyAlaProArgLysLysGlyCysTrpLysCySgIlySgIly 419  
QY 1246 GGCCACAGATGAAGACTGCACCGCAGCGCCAGCCCACTTCTGGGCAAGATCTGCGCC 1305  
Db 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439  
QY 1306 AGCCACAAGGCGCGCGCGCGCAAC----- 1329  
Db 440 SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459  
QY 1330 TTCTTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1383  
Db 460 PheLeuGlnSerArgProGluProThrAlaProProGluGlnSerPheArgSerGlyVal 479  
QY 1384 GAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437  
Db 480 GluThrThrThrProProGlnLysGlnGluProIleAspLysGluLeuTyrProLeuThr 499  
QY 1438 AGCCTGAAGAGCCTGTTGGCAACGACCGCCCTGAGCCAG 1476  
Db 500 SerLeuArgSerLeuPheGlyAsnAspProSerSerGln 512

RESULT 8

T09436  
gag polyprotein - human immunodeficiency virus type 1 (strain JRF1)  
C/Species: human immunodeficiency virus type 1, HIV-1  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C/Accession: T09436  
R/Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,  
submitted to the EMBL Data Library, July 1996  
A/Reference number: Z16673  
A/Accession: T09436  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-498 <PAN>  
A/Cross-references: UNIPROT:Q75754; EMBL:U63632; NID:g1465777; PID:g1465778  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polyprotein

Alignment Scores:

Pred. No.: 2.77e-106 Length: 498  
Score: 2233.00 Matches: 417  
Percent Similarity: 92.57% Conservative: 44

Best Local Similarity: 83.73% Mismatches: 31  
Query Match: 80.56% Indels: 6  
DB: 2 Gaps: 4

us-09-475-704a-3 (1-1479) x T09436 (1-498)

QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLysLeuAspLysTrpGluYsaIleArg 20  
QY 61 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
QY 121 CTGAGAAAGTTCGCGCTGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGlnSerSerGlnGlyCysArgGlnIle 60  
QY 181 ATCCGCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 61 LeuGlyGlnLeuLeuProAlaLeuLysThrGlySerGlnGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgGlnGlnValLysAspThrLysGlnAla 100  
QY 301 CTGCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuGluLysIleGluGluGluGlnAlaLysSerLysLysGluAlaAlaAlaAspThr 120  
QY 361 GCCGCAAGGCGAGGTGAGCGCAAGTACCCCATCGTGCGAAGCCTGCGCGCGCGAGATG 420  
Db 121 GlyAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnMetGlnGlyGlnMet 140  
QY 421 GTGCACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValValGluGluLys 160  
QY 481 GCCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Db 161 AlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThrProGln 180  
QY 541 GACCTGAACAGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
Db 181 AspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLys 200  
QY 601 GACACCATCAAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
Db 201 GluThrIleAsnGluGlnAlaAlaGluTrpAspArgLysHisProValHisAlaGlyPro 220  
QY 661 ATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
Db 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240  
QY 721 CTGCAGAGAGATCGCTGATGACCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
Db 241 LeuGlnGluGlnIleGlyTrpMetThrAsnAsnProProIleProValGlyGluIleTyr 260  
QY 781 AAGCGGTGATCATCTGCGGCTGAACAAGATCGTGCGGATGTACAGCGCGGTGAGCATC 840  
Db 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThrSerIle 280  
QY 841 CTGACATCAAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
Db 281 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPheTyrLys 300  
QY 901 ACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
Db 301 ThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGluThrLeuLeu 320  
QY 961 GTGCAGAGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020  
Db 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAlaAlaThr 340



DB 380 ArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGlnGly 399  
QY 1186 CACATCGCCCGCAACTGCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 1245  
DB 400 HsItleAlaLysAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGln 419  
QY 1246 GGCCACCAAGATGAAGAACTGCACCCGAGCGCCCAACTTCTGGGCAAGATCTGGCCC 1305  
DB 420 GlyHisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439  
QY 1306 AGCCACAAGGGCGCGCCCGCAACTTCTGCAGAGCGCGCCGAGCCCAACCGCCCCCCC 1365  
DB 440 SerHisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlnProThrAlaProSer 459  
QY 1366 GCGGAGAGCTTCCGCTTC-----GAGGAGACCACCCCGCGCAAGACGAGAGCAAG 1419  
DB 460 GlnGlnSerValArgPheGlyGlnGlnThrThrProSerGlnLysGlnGlnProIle 479  
QY 1420 GACCGCGAG-----ACCTTGACCAAGCTGAAGAGCTGTTCGGCAACGACCCCTGAGC 1473  
DB 480 AspLysGlnLeuTyrProLeuAlaSerLeuArgSerLeuPheGlySerAspProSerSer 499  
QY 1474 CAG 1476  
DB 500 Gln 500

RESULT 10  
S33979  
gag polyprotein - human immunodeficiency virus type 1  
C;Species: human immunodeficiency virus type 1, HIV-1  
C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
C;Accession: S33979  
R;Carlini, F.  
submitted to the EMBL Data Library, November 1991  
A;Reference number: S33979  
A;Accession: S33979  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-500 <CAR>  
A;Cross-references: UNIPROT:Q78240; EMBL:Z11530; NID:g60192; PIDN:CAA77621.1; PID:g60193  
C;Superfamily: AIDS-related virus gag polyprotein

Alignment Scores:  
Pred. No.: 4.97e-106 Length: 500  
Score: 2228.00 Matches: 418  
Percent Similarity: 91.62% Conservative: 41  
Best Local Similarity: 83.43% Mismatches: 32  
Query Match: 80.38% Indels: 10  
DB: 2 Gaps: 5

US-09-475-704A-3 (1-1479) x S33979 (1-500)

QY 1 ATGGGCGCCCGCGCGAGCATCTGCGCGCGCGAGCTGGAGCGCTGGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLysLeuAspArgTrpGlnLysIleArg 20  
QY 61 CTGGCGCCCGCGCGCAAGAAGTGTATCATGATGAGCACTGTGTGGCGCCAGCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysLysLysLysLysHisIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAGAAGTTGCGCTGAACCGCGCGCTGTGTGAGACCGAGCGGCTGCAAGCATC 180  
DB 41 LeuGlnArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60  
QY 181 ATCCGCCAGCTGACCGCGCGCTGTGAGACCGAGCGAGGAGCTGAAGAGCTGTCAAC 240  
DB 61 LeuGlnGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnArgArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCGCTGTACTGCGTGCAGAGAAGATGAGGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGlnIleLysAspThrLysGlnAla 100  
QY 301 CTGCAACAAGATCGAGAGAGAGAGACAAAGTGCACAGAGAAGATCCAGAGCGCGAGCC 360

DB 101 LeuAspLysIleGlnGlnGlnAlaSerLysLysLysAlaGlnGlnAla--Ala 119  
QY 361 GCCGACMAAGGC-----AAGGTGAGCCGAACCTACCCATCGTGCAAACTGACAG 411  
DB 120 AlaAspThrGlnHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139  
QY 412 GGCCAGATGTGCACCGCCATCAGCCCGCACCCCTGACCGCTGGGTGAAGGTGATC 471  
DB 140 GlyGlnMetValHisGlnAlaLeuSerProArgThrLeuAsnAlaTrpValLysVal 159  
QY 472 GAGGAGAAGGCTTCAAGCCGAGGTGATCCCATGTTCAACCGCTGAGGAGGGCGCC 531  
DB 160 GlnGlnLysAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnGlyAla 179  
QY 532 ACCCCCAAGACCTGAACACGATGTTGAACAAGTGGCGGCGCACAGCGCCATGACAG 591  
DB 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaIleMetGln 199  
QY 592 ATGTTGAAGACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGACCCCGTGAC 651  
DB 200 MetLeuLysGlnThrIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHisProValGln 219  
QY 652 GCCGCGCCCATCGCCCCCGCGCAGATGCGGAGCGCCCGCGCAGCAGCATCGCGCAC 711  
DB 220 AlaGlyProIleAlaProGlyGlnIleArgGlnProArgGlySerAspIleAlaGlyThr 239  
QY 712 ACCAGCACCTTGACGAGAGAGATCGCTGTGATGACCAACCCCGCATCCCGTGCGC 771  
DB 240 ThrSerThrLeuGlnGlnGlnIleArgTrpMetThrAsnAsnProProIleProValGly 259  
QY 772 GACATCTCAAGCGGTGATCATCTCGGCTGAACAAGATCGTGCGATGTACAGCCCC 831  
DB 260 GlnIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279  
QY 832 GTGAGCATCTTGACATCAAGCAGGCGCCCAAGAGCGCTTCGCGACTACGTGAGCCGC 891  
DB 280 ThrSerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArg 299  
QY 892 TTCTTCAAGACCTTCGCGCGCGAGCAGACCAAGAGGTGAAGAACTGATGACCGAC 951  
DB 300 PheTyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTrpMetThrGln 319  
QY 952 ACCCTGTGTGACAGAACCGCAACCCCGACTGCAAGACCATCTCGCGCTCGGCCCC 1011  
DB 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339  
QY 1012 GGCGCCAGCTTGAGAGATGATGACCGCTGCCAGGCGGTGGCGGCCCGACCAAG 1071  
DB 340 GlnAlaThrLeuGlnGlnGlnMetThrAlaCysGlnGlyValGlyGlyProGlnHisLys 359  
QY 1072 GCCCGGCTGTGCGCGAGCGGATGAGCCAGGCCAAC-----ACCAGCGTGTGATGACAG 1125  
DB 360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnProAlaAsnIleMetMetGln 379  
QY 1126 AAGAACAATTCAAGGGCGCGCGCGCATCGTCAAGTGTCTTCAACTGCGGCAAGAGGC 1185  
DB 380 ArgGlyLysPheArgAsnGlnGlyLysThrValLysCysPheAsnCysGlyLysGlnGly 399  
QY 1186 CACATCGCCCGCAACTGCCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGCAAGAG 1245  
DB 400 HsItleAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGln 419  
QY 1246 GGCCACCAAGATGAAGAACTGCACCCGAGCGCCCAACTTCTGGGCAAGATCTGGCCC 1305  
DB 420 GlyHisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439  
QY 1306 AGCCACAAGGGCGCGCCCGCAACTTCTGCAGAGCGCGCCGAGCCCAACCGCCCCCCC 1365  
DB 440 SerHisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlnProThrAlaProPro 459  
QY 1366 GCGGAGAGCTTCCGCTTC-----GAGGAGACCACCCCGCGCAAGAGAGAGCAAG 1419

```

Db      460  GluGluSerPheArgPheGlyGluGluThrThrProSerGlnGlnGluProIle 479
QY      1420  GACCCGGAG-----ACCTGACCAGCCTGAGAGCCTGTTCCGCAACGACCCCTGAGC 1473
      |||:::|||||  |||  |||:::|||||:::|||||:::|||||:::|||||
Db      480  AsplyGluMetTyrProLeuAlaSerLeuArgSerLeuPheGlyAsnAspProSerSer 499
QY      1474  CAG 1476
      |||
Db      500  Gln 500

RESULT 11
A38068
gag polyprotein - human immunodeficiency virus type 1 (strain MN)
N/Alternate names: core polyprotein
N/Contains: core protein p1; core protein p17; core protein p2; core protein p24; core p
C/Species: human immunodeficiency virus type 1, HIV-1
A/Note: host Homo sapiens (man)
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C/Accession: A38068
R./Henderson, L.B.; Bowers, M.A.; Sowder II, R.C.; Serabyn, S.A.; Johnson, D.G.; Bess Jr.
J. Virol. 66, 1856-1865, 1992
A./Title: Gag proteins of the highly replicative MN strain of human immunodeficiency viru
A./Reference number: A38068; MUID:92194415; PMID:1548743
A./Accession: A38068
A./Molecule type: protein
A./Residues: 1-506 <HEN>
A./Cross-references: UNIPROT:P05888
C./Genetics:
A./Gene: gag
C./Superfamily: AIDS-related virus gag polyprotein
C./Keywords: AIDS; core protein; immunodeficiency; polyprotein
F./1-134/Product: core protein p17 #status experimental <P17>
F./135-365/Product: core protein p24 #status experimental <P24>
F./366-379/Product: core protein p2 #status experimental <P2>
F./380-434/Product: core protein p1 #status experimental <P1>
F./435-450/Product: core protein p1 #status experimental <P1>
F./451-506/Product: core protein p6 #status experimental <P6>

Alignment Scores:
Pred. No.:      1  91e-105      Length:      506
Score:          2216.50      Matches:      418
Percent Similarity: 90.34%      Conservative: 40
Best Local Similarity: 82.45%      Mismatches: 32
Query Match:      79.96%      Indels:      17
DB:              1      Gaps:      5

US-09-475-704A-3 (1-1479) x A38068 (1-506)

QY      4  GGGGGCCGGCGGAGCATCTGCGGGCGGCGGAGCTGAGCGCTGGGAGCGCATCCGCTG 63
      |||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      1  GAlAAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTyrGluLysIleArgLeu 20
QY      64  CGCCCCGGCGGAGAAAGTGCTACATGATGAAGCACCTGTGTGGCCAGCCGCGAGCTG 123
      |||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      21  ArgProGlyGlyLysLysLysTyrLysLeuLysHisValIleTyrPalaSerArgGluLeu 40
QY      124  GAGAGTTGCGCCTGAAACCCCGGCTGCTGTGAGACCGAGGGGTGCAAGCAGATCATC 183
      |||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      41  GluArgPheAlaIleAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIleLeu 60
QY      184  CGCAGCTGCACCCCGCCCTGCGAGACCGGCGAGGAGGAGCTGAAGAGCCTGTTCAACACC 243
      |||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      61  GlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluArgLysSerLeuTyrAsnThr 80
QY      244  GTGGCCACCCCTGTAAGTGGTGCAGAGAAAGATCGAGGTCGGCGACCAAGAGAGCCCTG 303
      |||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      81  ValAlaThrLeuTyrCysValHisGlnLysIleLysIleLysAspThrLysGluAlaLeu 100
QY      304  GACAAAGATCGAGAGAGAGCAGAAACAGTGCCAGCAGAAAGATCCAGCAGGCCGAGCCGCC 363
      :::|||||:::|||||:::|||||:::|||||:::|||||
Db      101  GluLysIleGluGluGluGlnAsnLysSerLysLysAlaGlnGlnAla---AlaAla 119
      |||
364  GACAAAGGC-----AAGGTGAGCCAGAACTACCCCATCGTGCAAGAC 405

```

Db	120	A <del>sp</del> ThrGly <del>Asn</del> ArgGly <del>Asn</del> SerSerGlnValSerGlnAsnTyrProIleValGlnAsn	139
QY	406	CTGCAGGCGCCAGATGGTGACACGAGCCATCAGCCCCCGGACCTTGAACGCTGCGTGAAG	465
Db	140	IleGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLys	159
QY	466	GTGATCGAGGAGAAAGCGCTTCAGCCCCGAGGTGATCCCACTGTTCAACCGCCCTGAGCGAG	525
Db	160	ValValGlnGlnLysValAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGln	179
QY	526	GGCGGCCAACCCTCCAGGACCTGAACACGATGTTGAACACCGTGGCGGCGGACCGAGCGCGC	585
Db	180	GlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAla	199
QY	586	ATGCAGATGCTGAAGAACACCATCAACGAGAGAGCGCGCGAGTGGACCGCGTGACACCC	645
Db	200	MetGlnMetLeuLysGlnThrIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHisPro	219
QY	646	GTGCACGCGCGCCCATCGCCCCCGGCGGACGATGCGCGAGCCCCGCGGCGACATCGCC	705
Db	220	AlaHisValaglyProIleAlaProGlyGlnMetArgGlnProArgGlySerAspIleAla	239
QY	706	GGCACCAACGACACCCCTGCGAGAGAGATCGCTGATGACCAAGACCCCCCATCCCC	765
Db	240	GlyThrThrSerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsnProProIlePro	259
QY	766	GTGGCGGACATCTACAAGCGGTGGATCATCTGCGGCTGGAACAAGATCGTGGATGTAC	825
Db	260	ValGlyGlnIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyr	279
QY	826	AGCCCCGTGAGCATCCTGACATCAAGCAGAGGCCCCAAGAGCCCTTCGCGACTACGTG	885
Db	280	SerProSerSerIleLeuAsnProIleArgGlnGlyProLysGlnProPheArgAspTyrVal	299
QY	886	GACCGCTTCTTCAAGACCCCTGCGCGCGCGACGAGACACCAAGAGTGAAGACTGATG	945
Db	300	AspArgPheTyrLysThrLeuArgAlaGlnGlnAlaIleSerGlnGlnValLysAsnTrpMet	319
QY	946	ACCGACACCCCTGCTGTGTCGAGAAAGCCAAACCCCGACTGCAAGACCATCTGCGCGTCTC	1005
Db	320	ThrGlnThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeu	339
QY	1006	GGCCCCGGCGCCAGCCTGAGGAGATGATGACCGCCTGCCAGGCGTGGCGGCCCCAGC	1065
Db	340	GlyProAlaIaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGly	359
QY	1066	CACAAGCGCCCGCTGTGCGCGAGCGGATGACCGACGCGCAACACGAGC-----GTGATG	1119
Db	360	HisLysValaIaArgValIleuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMet	379
QY	1120	ATGCAGAAAGCAACTTCAAGGGCCCCCGCGCGCATCGTCAAGTGTCTTCAACTGCGGCAAG	1179
Db	380	MetGlnArgGlyAsnPheArgAsnGlnArgLysIleIleLysCysPheAsnCysGlyLys	399
QY	1180	GAGGGCCACATCGCCCGGCAACTGCGCGCGCCCCCGGACGAAGAAGGCTGCTGAAGTGGCGC	1239
Db	400	GlnGlyHisIleAlaLysAsnCysArgAlaProArgLysArgGlyCysTrpLysCysGly	419
QY	1240	AAGAGGGCCACCATGATGAGGACTGACCGGAGCGCGCAAGCCCACTTCTGGGCAAGATC	1299
Db	420	LysGlnGlyHisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIle	439
QY	1300	TGGCCCAAGCACAAGGGCGCCCCGGCAACTTCTGCAAGCGCGCCGAGCCACCGCC	1359
Db	440	TrpProSerCysLysGlyArgProGlnAsnPheProGlnSerArgThrGlnProThrAla	459
QY	1360	CCCCCGCGGAGAGCTTCGCGCTC-----GAGGAGACACCCCGCGCGCAAGACGAGAG	1413
Db	460	ProProGlnGlnLysPheArgPheGlyGlnGlnIleThrThrThrProTyrGlnLysGlnGln	479
QY	1414	-----AGCAAGAGACCGCGAGACCTTGACCAACGCTGGAAGAGCTGTTC	1455

Db 480 LysLysGlnGluThrIleAspLysAspLeuTyrProLeuAlaSerLeuLysSerLeuPhe 499  
QY 1456 GGCAACGACCCCTGAGCCAG 1476  
Db 500 GlyAsnAspProLeuSerGln 506

## RESULT 12

## FOVWVL

gag polyprotein - human immunodeficiency virus type 1 (isolate LV)

N;Alternate names: assemblin; core polyprotein; gag precursor

N;Contains: capsid antigen core protein p24CA; core protein p1; core protein p6; matrix

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004

C;Accession: A03948

R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.

Nature 313, 450-458, 1985

A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi

A;Reference number: A93355; MUID:8511157; PMID:2982104

A;Accession: A03948

A;Molecule type: DNA

A;Residues: 1-478 <MUE>

A;Cross-references: UNIPROT:P03350

C;Genetics:

A;Gene: gag

C;Superfamily: AIDS-related virus gag polyprotein

C;Keywords: AIDS; blocked amino end; core protein; immunodeficiency; lipoprotein; myrist

F;2-478/Product: gag precursor (assemblin) #status predicted <GAG>

F;2-132/Product: matrix antigen core protein p17MA #status predicted <P17>

F;20-32/Region: nuclear location signal

F;110-114/Region: nuclear location signal

F;133-363/Product: capsid antigen core protein p24CA #status predicted <P24>

F;364-377/Product: core protein p2 #status predicted <CP2>

F;378-432/Product: nucleocapsid core protein p7NC #status predicted <CP7>

F;392-405/Region: zinc finger CCHC motif

F;413-426/Region: zinc finger CCHC motif

F;2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

F;392,395,400,405/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

F;413,416,421,426/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

## Alignment Scores:

Pred. No.:	2,36e-96	Length:	478
Score:	2037.50	Matches:	388
Percent Similarity:	90.81%	Conservative:	47
Best Local Similarity:	81.00%	Mismatches:	33
Query Match:	73.50%	Indels:	11
DB:	1	Gaps:	8

US-09-475-704A-3 (1-1479) x FOVWVL (1-478)

QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGCGGCAAGCTGGACGCGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGCGGCAAGAGTGTCTATGATGATGAGCACTGTGTGGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
QY 121 CTGGAGAAGTTCGCGCTGAACCCCGCGCTGTGAGACCAAGGAGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60  
QY 181 ATCCGCGAGCTGCACCCCGCGCGCTGCAGACCGGAGCGAGGAGCTGAAGACCTGTTCAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCCCTGTACTGCGTGACAGAGAGATCGAGGTCCGCGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGlnIleLysAspThrLysGlnAla 100  
QY 301 CTGACACAGATCGAGAGGAGAGCAAGACAGTGCACAGAGATCCAGAGGCCGAGGCC 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysSerLysLysLysLysLysLysLysLysLys 119

QY 361 GCCGACAAAGGC-----AAGTGAGCCGAAGACTACCCCATCTGTGCAAACTTCAG 411  
Db 120 AlaSprThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGln 139  
QY 412 GGCCAGATGTGCACCAAGGCATCAGCCCGCGCACCCCTGAACGCTGGGTGAAGTGATC 471  
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValVal 159

QY 472 GAGGAGAAGGCTTCAGCCCGGAGGTGATTCCTCATGTCACCGCCCTGAGCGAGCGGCC 531  
Db 160 GluGlnLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179

QY 532 ACCCCCGAGACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCGCCCATGCAG 591  
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGln 199

QY 592 ATGCTGAAGACACCATCAAGAGAGGCGCGGAGTGGAGCCGCTGCACCCCGTGAC 651  
Db 200 MetLeuLysGluThrIleAsnGluGluAlaIleGluTrpAspArgValHisProValHis 219

QY 652 GCCGGCCCATCGCCCGCGCGAGTGCAGAGCCCGCGGAGCGGACATCGCCGCGAC 711  
Db 220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239

QY 712 ACCAGCACCTGCAGAGAGATGCGCTGATGACCAAGCAACCCCGCATCCCGTGCGGC 771  
Db 240 ThrSerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsnProProIleProValGly 259

QY 772 GACATCTACAAGCGGTGATCATCTGCGGCTGAACAAGATGTCGGATGATACAGCCC 831  
Db 260 GlnIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279

QY 832 GTGAGCATCTGGAACATCAAGAGAGGCGCCCAAGAGACCCCTTCGCGACTGATGAGCCC 891  
Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299

QY 892 TTCTTCAAGACCTTCGCGCGCGAGAGAGACCAAGAGGTGAAGAACTGTGATGACCGAC 951  
Db 300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGlyValLysAsnTrpMetThrGlu 319

QY 952 ACCCTGTGTGTCAGAAAGCGCAACCCCGACTGCAAGACCATCTGCGCGCTTCGCGCCC 1011  
Db 320 ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339

QY 1012 GGCGCGAGCCTTGAGAGATGATGACCGCTGCGCAAGGCGGTGGCGGCCGACCAAG 1071  
Db 340 AlaAlaThrLeuGluGluGlnMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359

QY 1072 GCCCGGTGTGCGCGCGAGGCGATGAGCCAG---GCCAACAC---AGCGTATGATGCAG 1125  
Db 360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln 379

QY 1126 AAGAGCAACTCAAGGGCGCGCGCGCATCTCAAGTGTCTCAACTGCGGCAAGAGGGC 1185  
Db 380 ArgGlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnCysGlyLysGluGly 399

QY 1186 CACATCGCCCGCAACTGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGCAAGAG 1245  
Db 400 HisThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGln 419

QY 1246 GGCCACAGATGAAGAGACTGCAGCGCGCGCAAGTCTGCGGCAAGATC--TGG 1302  
Db 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleCysLeu 439

QY 1303 CCCAGCCAAAGGCGCGCGCGCAACTTCTGTCAGAGCGCGCGCGAGCCGACCGCCCC 1362  
Db 440 ProThrArgGluGlyGln--GlyIlePhePhe-----ArgAlaAspGlnSerGlnGln 456

QY 1363 CCGCGCGAGAGCTTCGCTGAGAGAGACC---ACCCCGCGCGAGAGCAGAGAGAGC 1416  
Db 457 ProHisHisPhePheArgAlaAspGlnSerGlnGlnProHisGlnLysArgAlaSer 475

## RESULT 13

FOLJSI  
gag polypeptide - simian immunodeficiency virus SIVcpz  
N/Alternate names: core polypeptide  
C/Species: simian immunodeficiency virus SIVcpz  
A/Note: host Pan troglodytes (chimpanzee)  
C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C/Accession: S09983  
R/Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.  
Nature 345, 356-359, 1990  
A/Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.  
A/Reference number: S09983; MUID:90259077; PMID:2188136  
A/Accession: S09983  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-508 <HUE>  
A/Cross-references: UNIPROT:P17282; EMBL:X52154; NID:G58866; PIDN:CAA36401.1; PID:G58867  
C/Genetics:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polypeptide  
C/Keywords: AIDS; core protein; immunodeficiency; polypeptide

Alignment Scores:  
Pred. No.: 6.87e-94 Length: 508  
Score: 1989.00 Matches: 365  
Percent Similarity: 83.73% Conservative: 62  
Best Local Similarity: 71.57% Mismatches: 63  
Query Match: 71.75% Indels: 20  
DB: 1 Gaps: 4

US-09-475-704A-3 (1-1479) x FOLJSI (1-508)

QY 1 ATGGGCGCCCGCCGCGCATCTGCGCGCGCGCAAGCTTGAGCGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuThrGlyGlyLeuAspArgTrpGluValArg 20  
QY 61 CTGCGCGCCCGCGCGCAAGTGTCTACATGATGAAGCACTGTGTGGCGCGCGCGAG 120  
Db 21 LeuArgProGlyGlyArgGlyArgGlyMetCysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGATTGCGCTGGAACCCCGCGCTGTGAGACCAAGCGCGCTGCAAGCAGATC 180  
Db 41 LeuGluArgPheAlaCysAspProGlyLeuMetGlySerGlyGlyCysThrLeu 60  
QY 181 ATCCGCGCGCTGACCCCGCGCTGAGACCGCGCGAGAGAGCTGAAGCTGTTCAC 240  
Db 61 LeuGlnGlnLeuGlnProAlaLeuLeuThrGlySerGlyGlyLeuArgSerLeuPheAsn 80  
QY 241 ACCGTGCGCGCGCTGATCTGCTGCTGAGAGAGAGATGAGTCCGCGCGCAAGAGGCGC 300  
Db 81 ThrLeuAlaValLeuTrpCysHisLeuSerAspIleThrValGluAspThrGlnLeuAla 100  
QY 301 CTGACAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuGlnGlnLeuLeuArgHisHisGlyGlnGlnSerLeuThrGlySerAsnSerGly 120  
QY 361 GCCGACAGGCGCAAGGTGAGCCAG-----AACTAC 390  
Db 121 SerArgGlnGlyGlyAlaSerGlnGlyAlaSerAlaSerAlaGlyIleSerGlyAsnTrp 140  
QY 391 CCCATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 450  
Db 141 ProLeuValGlnAsnAlaGlnGlyMetValHisGlnAlaIleSerProArgThrLeu 160  
QY 451 AACGCTGGGTGAAGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 510  
Db 161 AsnAlaTrpValValValValGlnGlnValAlaPheSerProGlnValIleProMetPhe 180  
QY 511 ACCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570  
Db 181 SerAlaLeuSerGlnGlyAlaLeuProGlnAspValAsnThrMetLeuAlaValGly 200  
QY 571 GGCCACGAGCGCGCATGACAGATGCTGAAGAGACCACTCAAGAGAGAGCGCGAGTGG 630

Db 201 GlyHisGlnGlyAlaMetGlnValLeuGlyGluValIleAsnGlnGluAlaGluTrp 220  
QY 631 GACCGGTGACCCCGTGCAGCGCGCGCGCATGCGCGCGCGCGAGATGCGCGAGCGCGC 690  
Db 221 AspArgLeuHisProThrHisAlaGlyProIleAlaProGlyGlnLeuArgGluProArg 240  
QY 691 GGCAGGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750  
Db 241 GlySerAspIleAlaGlyThrThrSerThrLeuGlnGlnGlnIleGlyTrpThrAla 260  
QY 751 AACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810  
Db 261 AsnProProIleProValGlyAspValTrpArgGlyTrpValIleGlyLeuAsnLeu 280  
QY 811 ATCGTGGGATGTACAGCCCGTGCAGATCTGACATCAAGCAGAGCGCGCGCGCGCG 870  
Db 281 ValValArgMetCysProValSerIleLeuAspIleArgGlnGlyProGlyGluPro 300  
QY 871 TTCGCGACTACGTGACCGCGCTTTCAGAGACCTGCGCGCGCGCGCGCGCGCGCGCG 930  
Db 301 PheArgAspTrpValAlaAspArgPheTrpLeuArgAlaGlnAlaSerGlnGlu 320  
QY 931 GTGAAGACTGATGACCGACACCTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 990  
Db 321 ValLeuAsnTrpMetThrAspThrLeuLeuValGlnAsnAlaAsnProAspCysGln 340  
QY 991 ATCGTGGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1050  
Db 341 IleLeuLeuAlaLeuGlyProGlyAlaThrLeuGlnGlnMetMetThrAlaCysGlnGly 360  
QY 1051 GTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1098  
Db 361 ValGlyGlyProSerHisLeuAlaArgValLeuAlaGluAlaMetSerMetValGlnAsn 380  
QY 1099 CAGGCGCAACACCGCGTGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1158  
Db 381 GlnGlyArgAlaAspValPhePheGlnLeuGlyGlnGlyAlaGlyProLeuArgGlyIle 400  
QY 1159 AAGTCTTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1218  
Db 401 LeuCysPheAsnGlyGlyGlyGlnGlnHisLeuAlaArgAsnCysValAlaProArgArg 420  
QY 1219 AAGGCTGCTGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1278  
Db 421 LeuGlyCysTrpArgCysGlyGlnGlnGlnHisGlnMetLeuAspCysThrGlyArgGln 440  
QY 1279 GCCAATTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1338  
Db 441 ValAsnPheLeuGlyGlyGlyTrpProSerArgSerGlyArgProGlyAsnPheValGln 460  
QY 1339 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1398  
Db 461 AsnArgThrGlnProThrAlaProProIleGlnSerTrpGlyTrpGlnGlnGluGln 479  
QY 1399 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
Db 480 ---LysSerGlnGlnLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 498  
QY 1447 AGCTGTTCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476  
Db 499 SerLeuPheGlySerAspProSerSerGln 508

RESULT 14  
S08435  
gag polypeptide - human immunodeficiency virus type 2 D205  
C/Species: human immunodeficiency virus type 2 D205, HIV-2 D205  
C/Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004  
A/Accession: S08435  
R/Dietrich, U.; Adamski, M.; Kreutz, R.; Seipp, A.; Kuehn, H.; Ruebsamen-Waigmann, H.  
Nature 342, 948-950, 1989  
A/Title: A highly divergent HIV-2-related isolate.  
A/Reference number: S08434; MUID:90081881; PMID:2594088  
A/Accession: S08435



## Alignment Scores:

Pred. No.:	6.04e-64	Length:	521
Score:	1399.50	Matches:	285
Percent Similarity:	68.75%	Conservative:	78
Best Local Similarity:	53.98%	Mismatches:	119
Query Match:	50.49%	Indels:	46
DB:	2	Gaps:	13

US-09-475-704A-3 (1-1479) x SS3091 (1-521)

```
QY 1 ATGGGCGCGCGCGCAGATCCTGCGCGCGCAGCTGGAACGCTGGAGCGCATCCGC 60
Db 1 MetGlyAlaArgSerSerValLeuArgGlyLysLysAlaAspGluLeuGluLysValArg 20
QY 61 CTGCGCGCGCGCGCAGAGTGTACTATGATGAAGCACTGTGTGGCCAGCCGCGAG 120
Db 21 LeuArgProGlyLysLysLysLysTyrArgLeuLysHisValTyrAlaAlaAsnGlu 40
QY 121 CTGAGAGAGTTGCGCCCTGAACCCCGGCTGTGAGACCGAGCGAGGCTGCAAGCATC 180
Db 41 LeuAspArgPheGlyLeuAlaGluSerLeuLeuGluSerLysGluGlyCysGlnLysIle 60
QY 181 ATCCGCGAGCTGACCCCGCGCTGACAGCGGACGAGAGAGCTGAAGAGCTGTTCAC 240
Db 61 LeuThrValLeuAspProLeuValProThrGlySerGluAsnLeuLysSerLeuPheAsn 80
QY 241 ACCGTGGCCACCCCTGTACTGCGTGACAGAGAGATCGAGTCCGCGACACCAAGAGGCC 300
Db 81 ThrValCysValIlePheCysIleHisAlaGluGluValLysAspThrGluLysAla 100
QY 301 CTGACAAGATCGAGAGG-----AGCAGAACAGTGCCAGACAGAGA----- 343
Db 101 Lys-LysPheAlaArgArgHisIleValAlaGluThrGlyThrAlaGluLysMetProAs 120
QY 344 -TCCAGCAGGCGCGAGCGCGCGACAAAGGCGAGGTGAGCCAGAACTACCCCATCGTGAC 402
Db 120 nThrSerArgProThrAlaProProSerGlyArg--GlyGlyAsnTyrPro--ValGln 138
QY 403 AACGTGACGGGCGAGATGATGTCACCAAGCCATCAGCCCGCGACCCCTGAACGCTGGGTG 462
Db 139 AsnIleGlyLysAsnTyrThrHisValProLeuSerProArgThrLeuAsnAlaTyrVal 158
QY 463 AAGGTGATCGAGAGAGAGGCGCTTCAAGCCCGAGGTGATCCCATGTTCAACGCGCTGAGC 522
Db 159 LysLeuValGluGluLysLysPheGlyAlaGluValValProGlyPheGlnAlaLeuSer 178
QY 523 GAGGCGCGCACCCCGCGAGACCTGAACAGATGTTGAACACCGTGGCGCGCACAGGCC 582
Db 179 GluGlyCysThrProTyrAspIleAsnGlnMetLeuAsnCysValGlyAspHisGlnAla 198
QY 583 GCCATGACAGATGCTGAAGAGACACCATCAAGAGAGCGCGCGAGTGGGACCGCGTGCAC 642
Db 199 AlaMetGlnIleIleArgGluIleIleAsnGluAlaAlaAspTyrAspAlaGlnHis 218
QY 643 CCGGTGACGCGCGCGCCATCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGCATC 702
Db 219 ProIle---ProGlyProLeuProAlaGlyGlnLeuArgGluProArgGlySerAspIle 237
QY 703 GCGGCGACACACAGACCCCTGACAGAGACAGATCGCTGGATG---ACCAAGAACCCCGCC 759
Db 238 AlaGlyThrThrSerThrValAspGluGlnIleGlnTyrMetPheArgAlaGlnAsnPro 257
QY 760 ATCCCGGTGGGCGACATCTACAAGCGGTGATCATCTGGGCTGAACAAGATCGTGGC 819
Db 258 ValProValGlyAsnIleTyrArgArgTyrIleGlnIleGlyLeuGlnLysCysValArg 277
QY 820 ATGTACAGCCCGGTGACATCTTCAAGACCGTGGCGCGCGACAGAGAGCCCTTCCGCGAC 879
Db 278 MetTyrAsnProThrAsnIleLeuAspValLysGlnGlyProLysGluProPheGlnSer 297
QY 880 TACGTGACCGCTTCTTCAAGACCGTGGCGCGCGCGACAGACACCGAGGTGAAGAAC 939
Db 298 TyrValAspArgPheTyrLysSerLeuArgAlaGluGlnThrAspAlaAlaValLysAsn 317
```

```
QY 940 TGGATGACCGACACCCCTGTGTGTGACAGAACGCCAACCCCGACTGCAAGACCATCTGCGC 999
Db 318 TrpMetThrGlnThrLeuLeuIleGlnAsnAlaAsnProAspCysLysLeuValLeuLys 337
QY 1000 GCTCTCGGCGCGCGCGCGCGAGCTGAGAGAGATGATGACCGGCTGCCAGGGCGGTGGCGC 1059
Db 338 GlyLeuGlnMetAsnProThrLeuGluGluMetLeuThrAlaCysGlnGlyValGlyGly 357
QY 1060 CCCAGCCACAAGCGCGCGCTGTGCGCGAGCGGATGAGCCAGGCCAACACCAAGCTGATG 1119
Db 358 ProGlyGlnLysAlaArgLeuMetAlaGluAlaLeuLysGluAlaMetThrProAlaPro 377
QY 1120 ATGCAGAAAGCACTTCAAGGCGCGCGCGCGCATGCTCAAGTGTCTCAACTGCGCGCAAG 1179
Db 378 IleProPheAlaAlaAlaGln--GlnArgArgThrIleLysCysTyrPheAsnCysGlyLys 396
QY 1180 GAGGCGCACATCGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1239
Db 397 GluGlyHisSerAlaArgGlnCysArgAlaProArgArgGlnGlyCysTyrLysCysGly 416
QY 1240 AAGGAGGCGCACACAGATGAAGGATGCAACCGAGCGCGCGCGCAACTTCTGGGCAAGATC 1299
Db 417 LysThrGlyHisIleMetAlaLysCysProAspArgGlnAlaGlyPheLeuGly---Ile 435
QY 1300 TGGCGCGCGCACAAAGGCGCGCGCGCGCGCAACTTCTGCGAGCGCGCGCGAG----- 1350
Db 436 GlyProTyrGlyLys--LysProArgAsnPheProValAlaGlnValProGlnGlyLeu 454
QY 1351 ---CCACCGCGCGCGCGCGCGAG-----AGCTTCGCTTCGAGAGAGACCAACCCCGC 1398
Db 455 ThrProThrAlaProProAlaAspProAlaValAspLeuGluLysTyrMetGlnGln 474
QY 1399 GCGCAGAAAGCAG---GAGAGCAAGAGACCGC----- 1425
Db 475 GlyArgLysGlnArgGluGlnArgGluArgProTyrLysGluValThrGluAspLeuLeu 494
QY 1426 ----- 1446
Db 495 HisLeuGluGlnGlyLysThrProCysLysGluThrThrGluAspLeuHisLeuAsn 514
QY 1447 AGCTGTTCGGCAACGAC 1464
Db 515 SerLeuPheGlyThrAsp 520
```

Search completed: March 11, 2005, 15:43:00  
Job time : 87.7254 secs

**THIS PAGE BLANK (uspto)**

## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**

**THIS PAGE BLANK (USPTO)**



KW AIDS; Core protein; Polypeptide.  
SQ SEQUENCE 492 AA; 55085 MW; C4AF1F81CECEBD1D CRC64;

Alignment Scores:

Pred. No.:	1.14e-107	Length:	492
Score:	2606.00	Matches:	492
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	94.01%	Indels:	0
DB:	2	Gaps:	0

US-09-475-704A-3 (1-1479) x Q9WF75 (1-492)

```
QY 1 ATGGGCGCGCGCGCGCATCTCGCGCGCGCGAGCTGGACGCTGGAGCGCATCCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspAlaTrpGluArgIleArg 20
QY 61 CTGGCGCGCGCGCGCGAGAGTGTCTACATGATGACCACTGTGTGGCGCGCGAG 120
Db 21 LeuArgProGlyGlyLeuBlyScyBlyrMetMetLysHisLeuValTrpAlaSerArgGlu 40
QY 121 CTGAGAGAAGTTCGCGCTGAACCCCGCGCTGTGAGACCAAGCGGCTGCAAGACATC 180
Db 41 LeuGluLysPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
QY 181 ATCCGCGCAGCTGCACCCCGCGCGCGCGCGAGACCGGCGAGAGCTGAAGACCTGTTCAAC 240
Db 61 IleArgGlnLeuHisProAlaLeuGlnThrGlySerGluGluLeuLysSerLeuPheAsn 80
QY 241 ACCGTGGCGCGCGCTGTACTGCGTGCAGAGAGTCCGCGAGACCAAGAGAGGCC 300
Db 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100
QY 301 CTGACCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 101 LeuAspLysIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
QY 361 GCCGACCAAGGCGAGAGTGAAGCCAGAACTACCCCATCTGTGAGAACCTGAGGCGCAGATG 420
Db 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMet 140
QY 421 GTGCAACGAGCGCATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 160
QY 481 GCCTTCAGCGCGCGAGGTGATCCCATGTTCAACCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 161 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 180
QY 541 GACCTGAACACGATGTTGAACACCGGTGGCGCGCGCGCGCGCGCGCGCGCATGAGTCTGAAG 600
Db 181 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 200
QY 601 GACACCATCAACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 201 AspThrIleAsnGluGluAlaAlaGluTrpAspArgValHisProValHisAlaGlyPro 220
QY 661 ATCGCGCGCGCGCGAGATGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Db 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240
QY 721 CTGCAAGAGAGATCGCGCTGTGATGACCAAGCAACCCCGCATCCCGCTGGCGCGCATCTAC 780
Db 241 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyr 260
QY 781 AAGCGGTGATCATCTGGCGCTGAACCAAGATGTGCGGATGTACAGCCCGGTGAGCATC 840
Db 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 280
QY 841 CTGACATCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Db 281 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 300
```

```
QY 901 ACCCTGCGCGCGCGAGCAGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Db 301 ThrLeuArgAlaGluGlnSerThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 320
QY 961 GTGCAAGACCGCAACCCCGAGCTGCAAGACCACTCTGCGCGCTCTGCGCGCGCGCGAGC 1020
Db 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340
QY 1021 CTGAGAGAGATGATGACCGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGTG 1080
Db 341 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 360
QY 1081 CTGGCGCGAGCATGAGCCAGCGCAACCAAGCGTGTATGATGATGATGATGATGATGATG 1140
Db 361 LeuAlaGluAlaMetSerGlnAlaAsnThrSerValMetMetGlnLysSerAsnPheLys 380
QY 1141 GCGCGCGCGCGCGCATCGTCAAGTGTCTTCAACTGCGCGCAAGAGAGCGCACATCGCGCAAC 1200
Db 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlnHisIleAlaArgAsn 400
QY 1201 TGCGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisIleMetLys 420
QY 1261 GACTGCACCGAGCGCGCGCGCAACTTCTGCGCGCAAGATCTGCGCGCGCGCGCGCGCG 1320
Db 421 AspCysThrGluArgGlnAlaAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGCGCAACTTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 441 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 460
QY 1381 TTCGAGAGAGACCAAGCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 461 PheGluGluThrThrProGlyGlnLysGlnLysSerLysAspArgGluThrLeuThrSer 480
QY 1441 CTGAAGAGCGCTGTTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 481 LeuLysSerLeuPheGlyAsnAspProLeuSerGln 492
```

RESULT 2

Q9WF69 PRELIMINARY; PRT; 492 AA.

AC Q9WF69; 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Gag polypeptide.

GN Name=gag;

OS Human immunodeficiency virus 1.

OC Viruses; Retrovirdae; Lentivirus.

OX NCBI\_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99214383; PubMed=10196340;

RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,

RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,

RA Essex M.J.

RT "Molecular cloning and phylogenetic analysis of human immunodeficiency

RT virus type 1 subtype C: a set of 23 full-length clones from

RT Botswana."

RL J. Virol. 73:4427-4432(1999).

DR EMBL; AF110963; AAD17037.1; -.

DR PIR; S49086; S49086.

DR HSP; Q70622; IHVN.

DR GO; GO:0019012; C:virion; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016032; P:viral life cycle; IEA.

DR InterPro; IPR000721; Gag\_p24.

DR InterPro; IPR010999; Retrovir\_matrix.

DR InterPro; IPR000071; Retrovir\_p17.

DR InterPro; IPR008916; Retrov\_capsid\_C.

DR InterPro; IPR008919; Retrov\_capsid\_N.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PRINTS; PR00234; HIVMATRIX.  
 DR SMART; SM00343; Znf\_C2HC; 2.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 DR AIDS; Core protein; Metal-binding; Polyprotein; Zinc; Zinc-finger.  
 SQ SEQUENCE 492 AA; 55098 MW; C33DEDA2CB6ED800 CRC64;

## Alignment Scores:

Pred. No.:	2.32e-107	Length:	492
Score:	2599.00	Matches:	490
Percent Similarity:	99.80%	Conservative:	1
Best Local Similarity:	99.59%	Mismatches:	1
Query Match:	93.76%	Indels:	0
DB:	2	Gaps:	0

US-09-475-704A-3 (1-1479) x Q9WF69 (1-492)

QY 1 ATGGGCGCCGCGCGCATCTCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
 Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspAlaTrpGluArgIleArg 20  
 QY 61 CTGGCCCCGCGCGCAAGAAGTGTATGATGAAGCACTGTGTGGCCAGCCCGAG 120  
 Db 21 LeuArgProGlyGlyLysCysTyrMetMetLysHisLeuValTrpAlaSerArgGlu 40  
 QY 121 CTGGAGAAGTTGCGCTGAACCCCGCTGTGAGACAGCAGGCGCTGCAAGCATC 180  
 Db 41 LeuGluLysPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
 QY 181 ATCCGCGAGCTGACACCCCGCTGAGACCGGCAAGAGAGAGAGAGAGAGAGAG 240  
 Db 61 IleArgGlnLeuHisProAlaLeuGlnThrGlySerGluGluLeuLysSerLeuPheAsn 80  
 QY 241 ACCGTGGCCACCTGTACTGTGCTGAGCAGAGAAGATGAGTCCGCGCACACCAAGAGAG 300  
 Db 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100  
 QY 301 CTGGACAAGATCGAGAGAGAGAGACAAGTGCAGAGAGAGAGATCCAGAGCGCGAGCC 360  
 Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysIleGlnGlnAlaGluAla 120  
 QY 361 GCCGACAAGGCGCAAGTGAAGCCAGAACTACCCCATCGTCAGAACCTGCAGGGCCAGATG 420  
 Db 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 140  
 QY 421 GTGCACCAAGGCCATCAACCCCGCAGCCCTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480  
 Db 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 160  
 QY 481 GCCTTCAGCCCGGAGGTGATCCCATGTTCAACCCCTGAGCGAGGCGCCACCCCCAG 540  
 Db 161 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 180  
 QY 541 GACCTGAACACGATGTGAACACCGTGGGCGGCGCACAGCGCCCATGCAATGCTGAAG 600  
 Db 181 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 200  
 QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGCACCCCGTGCACGCGCGCCC 660  
 Db 201 AspThrIleAsnGluAlaAlaGluTrpAspArgValHisProValHisAlaGlyPro 220  
 QY 661 ATGCCCCCGGCGCAGATGCGCGAGCGCGCGCGGAGCGACATCGCGCGCACCAAGCACC 720  
 Db 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240  
 QY 721 CTGCAGAGAGAGATCGCTTGATGACCAAGCAACCCCCCATCCCGTGGGCGACATCTAC 780  
 Db 241 LeuGlnGluGlnIleAlaIleTrpMetThrSerAsnProIleProValGlyAspIleTyr 260

QY 781 AAGCGTGATCATCTGGGCTGAAACAGATCGTGCAGATGTACAGCCCGTGAGCATC 840  
 Db 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 280  
 QY 841 CTGCATCAAGAGGCGCCCAAGAGAGCCCTTCCGCACTACGTGAGCCGCTTCAAG 900  
 Db 281 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPheLys 300  
 QY 901 ACCGTGCGCGCGAGCAGAGACCCAGAGAGTGAAGAACTGATGACCGACACCTGCTG 960  
 Db 301 ThrLeuArgAlaGluGlnSerThrGlnGluValLysAsnTrpMetThrAspThrLeu 320  
 QY 961 GTGCAGAACGCCAACCCCGACCTGCAAGACCATCTGCGCGCTTCCGCGCGCGCAGC 1020  
 Db 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340  
 QY 1021 CTGCAGAGATGATGACCGCTGCGCAGGCGCTGGGCGGCGCCAGCCAGCGCGCTG 1080  
 Db 341 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 360  
 QY 1081 CTGGCCGAGCGCATGAGCCAGCGCAACACCATCGTATGATGACAGAAAGCACTTCAAG 1140  
 Db 361 LeuGlyGluAlaMetSerGlnAlaAsnThrAsnValMetMetGlnLysSerAsnLys 380  
 QY 1141 GGGCCCCGGCGCATCTGTCAAGTGTCTCAACTGCGCGCAAGAGGCGCACATCGCGCCAAC 1200  
 Db 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 400  
 QY 1201 TGCCCGCGCGCGCGCAAGAGGCTGTGAGAGTCCGCGCAAGAGAGGCGCACCATGTAAG 1260  
 Db 401 CysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGluGlyHisIleGlnMetLys 420  
 QY 1261 GACTGCACCGAGCGCCAGCGCACTTCTGTGGCAAGATCTGGCCCAAGCCACCAAGGCGCC 1320  
 Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
 QY 1321 CCGGCAACTTCTGTGACAGAGCGCGCGCGAGCCACCGCGCGCGCGAGAGCTTCCGC 1380  
 Db 441 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 460  
 QY 1381 TTGAGAGAGACCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 Db 461 PheGluGluThrThrProGlyGlnLysGlnLysSerLysAspArgGluThrLeuThrSer 480  
 QY 1441 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAG 1476  
 Db 481 LeuLysSerLeuPheGlyAsnAspProLeuSerGln 492

RESULT 3  
 Q9WF66 PRELIMINARY; PRT; 492 AA.  
 ID Q9WF66; AC Q9WF66; DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Gag polyprotein.  
 GN Name=gag;  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_Taxid=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vanberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makheima M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana."  
 RL J. Virol. 73:4427-4432(1999).  
 DR EMBL; AF110962; AAD17029.1; -.  
 DR HSSP; Q70622; 1HVN.

DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 492 AA; 55210 MW; 19657901783FEDBC CRC64;

## Alignment Scores:

Pred. No.:	1.64e-103	Length:	492
Score:	2512.00	Matches:	473
Percent Similarity:	98.37%	Conservative:	11
Best Local Similarity:	96.14%	Mismatches:	8
Query Match:	90.62%	Indels:	0
DB:	2	Gaps:	0

US-09-475-704A-3 (1-1479) x Q9WF66 (1-492)

QY 1 ATGGGCGCCGCGCCAGCATCTCTGCGCGCGCGCAAGCTGGACGCTGGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspAlaTrpGluLysIleArg 20  
QY 61 CTGCGCGCCGCGCGCAAGAAGTGTACATGATGAAGCACCTGTGTGGCCAGCCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysGlnTyrArgIleLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGAAGTTCGCTTGAACCCCGGCTGTGTGAGACCAGCGAGGCTGCAAGCATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATCCGCGAGCTGACCCCGCCTTGCAGACCGCGCGAGAGCTGAAGAGCTGTTCAC 240  
DB 61 IleGlnIleuGlnProAlaLeuGlnThrGlyThrGluIleuArgSerLeuPheAsn 80  
QY 241 ACCGTGGCCACCTGTACTGTGCTGTCAGAGAAAGATCGAGTCCGCGACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisLysLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTGCAGAGAAAGATCCAGAGCGCGAGCC 360  
DB 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysIleGlnIleGlnAla 120  
QY 361 GCCGACAAGGCGAAGGTGAGCGCAACTACCCCATCGTGCAACCTGCAGGGCGAGATG 420  
DB 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 140  
QY 421 GTGCACCAAGCCATCAGCCCGCGACCCCTGAACGCTGGTGAAGGTGATCGAGAGAAG 480  
DB 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 160  
QY 481 GCCTTCAGCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGGGCGCCACCCCCAG 540  
DB 161 AlaPheSerProGluValIlePrometPheThrAlaLeuSerGluGlyAlaThrProGln 180  
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGCGCCACCAAGCGCCATGCAGATGCTGAAG 600  
DB 181 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLys 200  
QY 601 GACACCATCAACAGAGAGCGCGCGAGTGGAGCGCGTGCACCCCGTGCACGCGCGCCCC 660  
DB 201 AspThrIleAsnGluGluAlaIleGluTrpAspArgLeuHisProValHisIleGlyPro 220  
QY 661 ATCGCCCCCGCGCAAGATGCGCGAGCCCGCGCGAGCAACATCGCCGCGCAACCAAGACC 720

DB 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240  
QY 721 CTGCAGAGACAGATCGCCTGGATGACCAGCAACCCCCCATCCCGGTGGCGCATCTAC 780  
DB 241 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyr 260  
QY 781 AAGCGGTGATCATCTCTGGGCTGAACAAGATCGTGCAGATGTACAGCCCCGTGAGCATC 840  
DB 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 280  
QY 841 CTGCACATCAAGCAGAGGCGCCCAAGAGCGCCTTCCGCGACTACGTGAGCCGCTTCTCAAG 900  
DB 281 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 300  
QY 901 ACCCTGCGCGCGAGCAGAGACCCAGAGGTGAAGAACTGATGACCCGACCCCTGCTG 960  
DB 301 ThrLeuArgAlaGluGlnSerThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 320  
QY 961 GTGCAGAACCGCCAACCCCGACTGCAGAACCATCTGCGCGCTCTCGGCCCGCGCGCAGC 1020  
DB 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340  
QY 1021 CTGAGAGAGATGATGACCGCTTGCAGAGGCGGTGGCGCGCCCAAGCAGAGCCCGCTG 1080  
DB 341 LeuGluGlnMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 360  
QY 1081 CTGGCGCGAGCGATGAGCCGCGCAACCAACCAACCGTGTATGATGCAAGAAGCAACTTCAAG 1140  
DB 361 LeuAlaGluAlaMetSerGlnThrAsnThrAsnValMetMetGlnArgSerAsnPheLys 380  
QY 1141 GCGCCCGCGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGAGGCGCACATCGCCCGCAAC 1200  
DB 381 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 400  
QY 1201 TGCCGCGCGCGCGCAAGAAGGCTGTGTGAAGTGCAGAGAGAGGCGCCACCAAGATGAAG 1260  
DB 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleGlnMetLys 420  
QY 1261 GACTGCACCGAGCGCCAGGCGCAACTCTCTGGGCAAGATCTGGCCCAAGGCGCGCC 1320  
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
QY 1321 CCGGCAACTTCTTGCAAGAGCGCGCGCGCGCCAGCCCGCGCGCGAGAGCTTCCGC 1380  
DB 441 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 460  
QY 1381 TTGAGAGAGACCAACCCCGCGCGCAAGAGCAGAGAGCAAGGACCGGAGACCTTGACGAC 1440  
DB 461 PheGluGluThrThrProValProLysGlnGluProLysAspArgGluProLeuThrSer 480  
QY 1441 CTGAAGAGCTGTTGGCAACGACCCCTGAGCCAG 1476  
DB 481 LeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 4  
Q9WF72 PRELIMINARY; PRT; 492 AA.  
ID Q9WF72;  
AC Q9WF72;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Gag polyprotein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;

RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana."  
RL J. Virol. 73:4427-4432(1999).  
DR EMBL; AF110964; AAD17046.1; -.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:viral ion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVIMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.  
SQ SEQUENCE 492 AA; 55329 MW; 56A4D9E3457F84C5 CRC64;

Alignment Scores:

Pred. No.:	1.64e-103	Length:	492
Score:	2512.00	Matches:	471
Percent Similarity:	98.58%	Conservative:	14
Best Local Similarity:	95.73%	Mismatches:	7
Query Match:	90.62%	Indels:	0
DB:	2	Gaps:	0

US-09-475-704A-3 (1-1479) x Q9WF72 (1-492)

QY 1 ATGGGCGCCGCGCGCATCTCTGCGCGCGCAAGCTGAGCGCTGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspAlaTrpIleArg 20  
QY 61 CTGGCGCCCGCGCGCAAGTGTACATGATGAAGCACTGTGTGGCCAGCGCGAG 120  
DB 21 LeuArgProGlyGlyLeuArgGlyGlyLeuArgGlyGlyLeuValTrpAlaSerArgIle 40  
QY 121 CTGGAAGATTGCGCTGAACCCGCGCTGCTGAGACGAGCGGCTGCAAGCATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrLeuGluCysIle 60  
QY 181 ATCGCGCAGCTGACCCCGCGCTGACGAGCGAGCGAGCTGAAGACCTGTCAAC 240  
DB 61 IleGlnIleuGlnProAlaLeuGlnThrGlyThrGluIleuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisIleArgIleGluIleArgAspThrIleGluAla 100  
QY 301 CTGACCAAGATCGAGAGAGAGACAGACAAGTCCGAGAGAGATCCAGCAGCGCGGCC 360  
DB 101 LeuAspIleIleGluGluGlnAsnIleCysGlnIleIleGlnIleAlaGluAla 120  
QY 361 GCCGACCAAGGCGAGGTGAGCCAGAACTACCCCATGTGCAAGACCTGCAAGGCGCAGATG 420  
DB 121 AlaAspIleGlyLeuValSerGlnAsnTyrProIleValGlnAsnLeuGlnIleMet 140  
QY 421 GTGACCAAGGCGCATAGCCCGCGACCTGAACGCTGGGTGAAGGTGATCGAGAGAAAG 480  
DB 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValIleGluGluIle 160  
QY 481 GCCTTCAAGCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCCAG 540  
DB 161 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 180  
QY 541 GACCTGAACAAGATGTGAACAACCGTGGCGCGCCACCAAGCGCCCATGACATGCTGAAG 600

DB 181 AspleuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuIys 200  
QY 601 GACACCATCAACGAGAGAGCGCCGAGTGGAGCCGCGTGCAACCCCGTGACGCGGCCCC 660  
DB 201 AspThrIleAsnGluIleAlaIleAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 220  
QY 661 ATCGCCCCCGCGCATGCGCGAGAGCGCCGCGCGAGCATCGCGCGCACCAAGACACC 720  
DB 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240  
QY 721 CTGACGAGACAGATCGCTGATGACCAACACCCCCCATCCCGTGCGGCACTTCTAC 780  
DB 241 LeuGlnIleGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyr 260  
QY 781 AAGCGGTGATCATCTCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCGTGACATC 840  
DB 261 LysArgTrpIleIleLeuGlyLeuAsnIleValAlaArgMetTyrSerProValSerIle 280  
QY 841 CTGACATCAAGCAGAGGCGCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 900  
DB 281 LeuAspIleIleGlnIleGlyProIleGlyProPheArgAspTyrValAspArgPheIle 300  
QY 901 ACCGTGGCGCGCGAGACAGACACCAAGAGGTGAAGAACTGATGACCGACACCTGCTG 960  
DB 301 ThrLeuArgAlaGluGlnIleThrGlnIleValIleAsnTrpMetThrAspThrLeuLeu 320  
QY 961 GTGACAGACGCGCAACCCCGACTGCAAGACCATCTGCGCGCTTGGGCGCGCGCAGC 1020  
DB 321 ValGlnAsnAlaAsnProAspCysIleThrIleLeuArgAlaLeuGlyProGlyAlaSer 340  
QY 1021 CTGAGAGAGATGATGACCGCTGCGCAGGCGTGGCGCGCGCCAGCACAAGCGCGGTG 1080  
DB 341 LeuGluGluMetMetThrAlaCysGlnIleValGlyGlyProSerHisIleValAlaArgVal 360  
QY 1081 CTGGCGCGAGCGGATGAGCCAGGCGCAACACCGGTGATGATGCAAGAGCAACTTCAAG 1140  
DB 361 LeuAlaGluAlaMetSerGlnThrAsnThrAsnValMetMetGlnArgSerAsnIleIys 380  
QY 1141 GCGCCCGCGCGCATGCTCAAGTGTCTCAACTGCGCGCAAGAGGCGCACATGCGCGCAAC 1200  
DB 381 GlyProArgArgIleValIleValIleCysPheAsnIleCysGlyIleGluGlnIleHisIleAlaArgAsn 400  
QY 1201 TGGCGCGCGCGCGCAAGAGGCTGCTGAACTGCGCGCAAGAGGCGCACATGCGCGCAAG 1260  
DB 401 CysArgAlaProArgIleValIleValIleCysPheAsnIleCysGlyIleGluGlnIleHisIleAlaArgAsn 420  
QY 1261 GACTGCACCGAGCGCGAGCGCAACTTCTGGGCAAGATCTGGCCAGCCACCAAGGCGCGC 1320  
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyIleTrpProSerHisIleValArg 440  
QY 1321 CCGGCAACTTCTGACAGAGCGCGCGAGCGCACCGCGCGCGCGAGAGCTTCCGC 1380  
DB 441 ProGlyAsnPheLeuGlnIleSerArgProGluProThrAlaProProAlaGluSerPheArg 460  
QY 1381 TTGAGAGAGACCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 461 PheGluGluThrThrProValGlnIleGlnIleProIleAspArgGluThrLeuThrSer 480  
QY 1441 CTGAAGAGCCTGTTCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476  
DB 481 LeuIleSerLeuPheGlySerAspProLeuSerGln 492  
RESULT 5  
Q8UT62 ID PRELIMINARY; PRT; 492 AA.  
AC Q8UT62;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxId=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21988475; PubMed=11991972;  
RX DOI=10.1128/JVI.76.11.5435-5451.2002;  
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,  
RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,  
RA Foley B.T., Gaoekwe S., Rybak N., Gasettsiwe S., Vanberg F.,  
RA Marlink R., Lee T.-H., Essex M.;  
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:  
RT consensus sequence for an AIDS vaccine design?";  
RL J. Virol. 76:5435-5451(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,  
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gasettsiwe S., Vanberg F.,  
RA Marlink R., Lee T.-H., Essex M.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF443100; AAL34787.1; -.  
DR HSSP; P24736; INCP.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf\_CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
DR AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 492 AA; 55149 MW; 61B3F3B5E87B9A1B CRC64;

Alignment Scores:  
Pred. No.: 3.62e-101 Length: 492  
Score: 2459.00 Matches: 462  
Percent Similarity: 97.36% Conservative: 18  
Best Local Similarity: 93.71% Mismatches: 11  
Query Match: 88.71% Indels: 2  
DB: 2 Gaps: 2

US-09-475-704A-3 (1-1479) x Q8UT62 (1-492)

QY 1 ATGGGGCGCCGCGCGCATCTCTGGCGCGGCGCAAGCTGGACGCTGGAGCGCATCCGC 60  
DB 1 Metc1yAlaArg1aSer1leLeuArgGlyGlyLysLeuAspLysTrpGluArgIleArg 20  
QY 61 CTGGCGCCCGCGCGCAAGAGTGTATGATGTAAGCACTGTGTGGCCAGCGCGGAG 120  
DB 21 LeuArgProGlyGlyLysLysGlnTyrMetIleLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGGAGAAGTTCGCCCTGAACCCCGGCTGTGGAGACCGAGGAGCTGCAAGAGATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrAlaGluGlyCysLysGlnIle 60  
QY 181 ATCCGCCAGCTGCACCCCGCCCTGTGACACCGGCAAGAGTGAAGAGCTGTTCAC 240  
DB 61 IleArgGlnLeuHisProAlaLeuGlnThrGlyThrGlnGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTCTGTACTGCGGTGACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisLysAspIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGACAAGATCGAGGAGGAGGACAGAACAGTGCAGCAAGAGATCGACAGGCGGAGCC 360  
DB 101 LeuAspLysIleGluGlnGluGlnAsnLysSerGlnGlnLysAlaGlnAlaGluAla 120  
QY 361 GCCGACAAGGGCAAGGTAGCCAGAACTACCCCATCGTGCAAGAACTGCAGGGCCAGATG 420  
DB 121 AlaAspLys--GlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 139  
QY 421 GTGCAACGAGGCCATCAGCCCGGACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480

DB 140 ValHisGlnProLeuSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159  
QY 481 GCCTCAGCCCGGAGGTGATCCCATGTTACCGCCCTGAGCGAGGCGCCACCCCGAG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179  
QY 541 GACCTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGACAGTGTGAAG 600  
DB 180 AspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAACGAGAGGAGCGCGCGAGTGGACCGCGTGCACCCCGTGCACGCGCGCCC 660  
DB 200 AspThrIleAsnGluGluAlaIleGluThrAspArgLysHisProValHisIleGlyPro 219  
QY 661 ATGCCCCCGGCGAGATGCGCGGAGCCCGCGGCGAGACATCGCCGCGCACAGCACCC 720  
DB 220 ValAlaProGlyGlnMetArgAspProArgLysSerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGCAGAGCAGATCGCCTGATGACCAAGACCCCGCATCCCGTGGCGGACATCTAC 780  
DB 240 LeuGlnGluGlnIleThrTrpMetThrSerAsnProProIleProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGGCGCTGAACAAGATCGTGCGATGTACAGCCCGGTGACATC 840  
DB 260 LysArgTrpIleValLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279  
QY 841 CTGCACATCAACGAGGCGCCCAAGAGGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900  
DB 280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGGAGCAGAGCAACCCAGAGGTGAAGAACTGGATGACCGACACCTGTCTG 960  
DB 300 ThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAGAACGCCAACCCCGACTGCAAGACCAATCTGCGCGCTTCTGGCCCGCGCCAGC 1020  
DB 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 339  
QY 1021 CTGGAGAGATGATGACCGCGCTGCGAGGCGGTGGCGGCGCCAGCCACAAGCGCGCTG 1080  
DB 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359  
QY 1081 CTGGCCGAGGCGATGAGCCAGGCC--AACACAGCGTGTATGACAGAGCAACTTC 1137  
DB 360 LeuAlaGluAlaMetSerGlnAlaThrAsnThrSerIleMetMetGlnLysSerAsnPhe 379  
QY 1138 AAGGCGCCCGCGCGCATGCTGAAGTGTCTTAACTGCGGCAAGAGGCGCCACATCGCCGC 1197  
DB 380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlnHisIleAlaLys 399  
QY 1198 AACTGCGCGCGCCCGGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGCCACAGATG 1257  
DB 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
QY 1258 AAGGACTGACGAGCGCGCCAGGCCAACTTCTGGGCAAGATGTGCCCAAGAGGC 1317  
DB 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439  
QY 1318 CGCCCGGCAACTTCTGACAGCGCGCCCGGAGCCACCGCGCCCGCGCGAGAGCTTC 1377  
DB 440 ArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGlnSerPhe 459  
QY 1378 CGCTTGAGAGACCAACCCCGGCGCAGAGCGAGAGAGAGAGACCGCGAGCCCTGAGCC 1437  
DB 460 LysPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuThr 479  
QY 1438 AGCCTGAAGAGCCTGTTGGCAAGAGACCGCCCTGAGCCAG 1476  
DB 480 SerLeuLysSerLeuPheGlyAsnAspProLeuSerGln 492

RESULT 6  
Q8AFG0 ID Q8AFG0 PRELIMINARY; PRT; 493 AA.

AC Q8AFG0; (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, last annotation update)  
DE Gag protein.  
OS Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=98ZA445;  
RX MEDLINE=22567157; PubMed=12680693; DOI=10.1023/A:1022376022104;  
RA Hunt G.M., Papathanasopoulos M.A., Gray G.E., Tiemeissen C.T.;  
RT "Characterisation of near-full length genome sequences of three South  
African human immunodeficiency virus type 1 subtype C isolates."  
RL Virus Genes 26:49-56(2003).  
DR EMBL; AY158533; AAN77379.1; -.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:viralion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf\_CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 493 AA; 55273 MW; DB51A1997BC32EDD CRC64;

Alignment Scores:  
Pred. No.: 5.16e-101 Length: 493  
Score: 2455.50 Matches: 459  
Percent Similarity: 97.16% Conservative: 20  
Best Local Similarity: 93.10% Mismatches: 13  
Query Match: 88.58% Indels: 1  
DB: 2 Gaps: 1

US-09-475-704A-3 (1-1479) x Q8AFG0 (1-493)

QY 1 ATGGGCGCGCGCGCGCATCTCGCGCGCGCGCAAGCTGGACGCGCTGGGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspThrTrpGluLysIleArg 20  
QY 61 CTGGGCGCGCGCGCGCAAGAGTGTCTATGATGAGCACTGTGTGGGCGAGCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysIleTyrMetIleLysIleValIleValIleValIleVal 40  
QY 121 CTGGAGAAGTTCGCGCTGAACCGCGCGCTGTGAGACCGAGCGGCTGCAAGCAGATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATCCGCCAGCTGCACCGCGCGCTGAGACCGGCGAGGAGCTGAGAGCCTGTTCAAC 240  
DB 61 LeuLysGlnLeuHisProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGTGACGAGAGATCGAGGTCCGCGACACCAAGAGAGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisGluArgIleGluIleArgAspThrLysGlnAla 100  
QY 301 CTGGACAAGATCGAGAGAGAGAGAGAGAGAGAGATCGAGAGATCGAGAGAGAGAGAGCC 360  
DB 101 LeuAspLysIleGluGluGluGluGlnAsnLysValGlnGlnLysAlaGlnAlaGlnAla 120  
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGCGAGGCGCGAGATG 420

DB 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnIleGlnGlyGlnMet 140  
QY 421 GTGCACCAAGCCATACGCCCCCGCACCTTGAAAGCGCTGGTGAAGGTGATCGAGAGAGAG 480  
DB 141 ValHisGlnProLeuSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLys 160  
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCGCTGAGCGAGGCGCCACCCCCAG 540  
DB 161 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 180  
QY 541 GACCTGAACACGATGTTGAACACCGGTGGCGCGCGCACGACATGCGCGGACACGACGACC 600  
DB 181 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 200  
QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCCGGTGCACCCCGTGCACGCGCGGCC 660  
DB 201 AspThrIleAsnGluGlnAlaAlaGluThrAspArgLeuHisProValHisAlaGlyPro 220  
QY 661 ATCGCCCCCGCGCAGATGCGCGAGCGCGCGCGCGCGAGACATGCGCGGACACGACGACC 720  
DB 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerAsn 240  
QY 721 CTGAGAGAGCAGATCGCCTGTGATGACGACCAACCCCATCCCGGTGGCGACATCTAC 780  
DB 241 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyr 260  
QY 781 AACCGGTGATCATCTCGGCGCTGGAACAAGATCGTCGGATGTACAGCCCGTGAGATC 840  
DB 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 280  
QY 841 CTGACATCAAGCAGGCGCGCGCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900  
DB 281 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 300  
QY 901 ACCGTGCGCGCGCGAGCAGACCAAGAGAGGTGAAGAACTGATGACCGACACCTGTCTG 960  
DB 301 ThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 320  
QY 961 GTGCAGAAGCCCAACCCCGACTGCAAGACCATCTGCGCGCTTCGCGCGCGCGCGCAGC 1020  
DB 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyLysSer 340  
QY 1021 CTGAGAGAGATGACCGCGCTGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCTG 1080  
DB 341 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysValAlaArgVal 360  
QY 1081 CTGGCGGAGCGGATGAGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137  
DB 361 LeuAlaGluAlaMetSerGlnAlaGlyAsnSerAsnIleMetMetGlnLysSerAsnPhe 380  
QY 1138 AAGGCGCGCGCGCGCATCGTCAAGTGTCTCAACTCGCGCGCAAGAGGCGCGACATCGCCGC 1197  
DB 381 LysGlyProArgArgThrValLysCysPheAsnLysGlyLysGluGlnHisIleAlaLys 400  
QY 1198 AACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1257  
DB 401 AsnCysArgAlaProArgLysLysGlyCysTrpAsnLysGlyLysGluGlnHisIleGlnMet 420  
QY 1258 AAGACTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1317  
DB 421 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 440  
QY 1318 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1377  
DB 441 ArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPhe 460  
QY 1378 CGCTTCGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437  
DB 461 ArgPheGluGluAspThrProValProLysGlnGluGlnLysGluArgGluProLeuThr 480  
QY 1438 AGCCTGAAGAGCCTGTTCCGCAACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1476  
DB 481 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 493



Db 460 ArgPhegluGlutThrProAlaProlysGlnGluProileGluArgGluProleuThr 479

QY 1438 AGCCTGAAGAGCCTGTTGGCAACGACCCCTGAGCCAG 1476

Db 480 SerleuLysSerleuPheGlySerAspProleuSerGln 492

RESULT 8

Q8UT89 ID Q8UT89 PRELIMINARY; PRT; 492 AA.

AC Q8UT89; 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE Gag protein.

GN Name=gag;

OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21988475; PubMed=11991972;

RX DOI=10.1128/JVI.76.11.5435-5451.2002;

RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P., Williams C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I., Foley B.T., GaoJekwe S., Rybak N., Gaseitsiwe S., Vamberg F., Marlink R., Lee T.-H., Essex M.;

RA "Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design?";

RT J. Virol. 76:5435-5451(2002).

RL [2]

RN SEQUENCE FROM N.A.

RP Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vamberg F., Marlink R., Lee T.-H., Essex M.;

RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

RL EMBL; AF443097; AAL34760.1; -.

DR HSSP; P24736; INCP.

DR GO; GO:0019012; C:virion; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016032; P:viral life cycle; IEA.

DR Pfam; PF00540; Gag\_p17; 1.

DR Pfam; PF00607; Gag\_p24; 1.

DR Pfam; PF00098; zf\_CCHC; 2.

DR SMART; SM00343; Znf\_C2HC; 2.

DR PROSITE; PS50158; zf\_CCHC; 2.

DR AIDS; Core protein; Polyprotein.

SQ SEQUENCE 492 AA; 5516 MW; 82973B40D175A5D4 CRC64;

Alignment Scores:

Pred. No.: 3.4e-100 Length: 492

Score: 2437.00 Matches: 457

Percent Similarity: 97.16% Conservative: 22

Best Local Similarity: 92.70% Mismatches: 12

Query Match: 87.91% Indels: 2

DB: 2 Gaps: 2

US-09-475-704A-3 (1-1479) x Q8UT89 (1-492)

QY 1 ATGGGCGCCCGCCGACATCTCGCGCGCGGCAAGCTGGACGCCCTGGAGCGCATCCGC 60

Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGluLysIleArg 20

QY 61 CTGGCGCCCGCGGCAAGAGTGCCTACATGATGAAGCACTGTGTGGGCCAGCCGCGAG 120

Db 21 LeuArgProGlyGlyLysLysHisTrpMetMetLysHisIleValTrpAlaSerArgL 40

QY 121 CTGGAAGAAGTTCGCTCGAAGCCCGGCTGTGAGACCAAGGGGCTGCAAGCAGATC 180

Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60

QY 181 ATCCGCCAGCTGCACCCCGCCCTGCAGACCGGCAAGGAGAGCTGAAGCCTGTTCAAC 240

Db 61 MetLysGlnLeuGlnProAlaIleGlnThrGlyThrGluGluLeuArgSerLeuPheAsn 80

QY 241 ACCGTGCCACCCCTGTACTGCGTGCACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300

Db 81 ThrValAlaThrLeuTyrcysValHisGluGlyLeaAspValGlnAspThrLysGlnAla 100

QY 301 CTGACAAGATCGAGAGAGAGACAAGTGCCAGACAGAAAGATCCAGAGCCGAGGCC 360

Db 101 LeuAspLysValGluGluGlnAsnLysSerGlnGlnLysMetGlnAlaGlnAla 120

QY 361 GCCGACAAGGCGAAGGTGAGCGCAACTACCCCATCGTGACAACTGCAGGGCCAGATG 420

Db 121 AlaAspLys--LysValSerGlnAsnTyrcysValGlnValGlnAsnLeuGlnGlnMet 139

QY 421 GTGCAACGAGCCATCAGCCCGCACCTGAAACGCTGGTGAAGTGATCGAGAGAAG 480

Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluAspLys 159

QY 481 GCCTTCAGCCCGGAGTGATCCCATGTTCAACGCGCTGAGCGAGGGCCACCCCCAG 540

Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179

QY 541 GACCTGAACAGATGTTGAACACCGTGGCGGCGCACAGCCGCGCATGACATGCTGAAG 600

Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199

QY 601 GACACCATCAACGAGAGAGCGCGGAGTGGAACCGCGTGACACCCCGTGACAGCCGCGCC 660

Db 200 AspThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisProValHisAlaGlyPro 219

QY 661 ATCGCCCGCGCGGAGATGCGCGAGCCCGCGGCGACGACATCGCGCGCACCAAGACAC 720

Db 220 ValAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239

QY 721 CTGCAAGAGCAGATCGCTGTGATGACCAAGACCCCCCATCCCGGTGGCGACATCTAC 780

Db 240 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTy 259

QY 781 AAGCGGTGATCATCTTGGGCTGAAACAAGATCGTGGGATGTACAGCCCGTGACATC 840

Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrcysProValSerIle 279

QY 841 CTGACATCAAGAGGGGCGCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAG 900

Db 280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyValAspArgPheNhelys 299

QY 901 ACCCTGCGCGCGAGCAGACAGACCCAGAGGTGAAGAACTGGATGACCGACCCCTGCTG 960

Db 300 ThrLeuArgAlaGluGlnSerThrGlnGluValLysAsnTrpMetThrAspThrLeuLeu 319

QY 961 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTTGGCGCTTCTGGCCCGCGCGCAGC 1020

Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlySerGlyAlaSer 339

QY 1021 CTGAGAGAGATGATGACCGCTGCGCAGGGCGGTGGCGGCGCCAGCACAAGGCCGCGTG 1080

Db 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359

QY 1081 CTGGCCGAGGCGATGAGCCAGGCC--AACACCAAGCTGATGATGACAGAAGACAATTTC 1137

Db 360 LeuAlaGluAlaMetSerGlnAlaAsnValGlnIleMetMetGlnLysAsnAsnPhe 379

QY 1138 AAGGGCCCGCGCGCATGCTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCGC 1197

Db 380 LysGlyProArgArgIleIleLysCysPheAsnCyGlyLysGluGlnHisLeuAlaArg 399

QY 1198 AACTGCGCGCGCCCGCAAGAGGGCTGTGAAGTGCGGCAAGAGGGCCACAGATG 1257

Db 400 AsnCyAspArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisGlnMet 419

QY 1258 AAGCACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCAGCCACAAGGGC 1317

DB 420 LysaspCysThrGluArgGlnAlaAsnPhleuGlyLysIleTrrPProserHisLysGly 439  
QY 1318 CGCCCCGGCACTTCTGACAGACCGCCCCGAGCCACCGCCCCCGCGAGAGCTTC 1377  
DB 440 ArgProGlyAsnPhleuGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTCGAGGAGACCAACCCCGCGCAGAGAGCAAGCAAGCCGCGAGACCTGACC 1437  
DB 460 ArgPheGluGluThrThrProserProLysGlnGluLeuLysAspLysGluProLeuThr 479  
QY 1438 AGCTGAAGAGCCTGTTCGCAACGACCCCTGAGCCAG 1476  
DB 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 9

Q901X0 PRELIMINARY; PRT; 492 AA.  
AC Q901X0;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_Taxid=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21565558; PubMed=11709097; DOI=10.1089/08892220152644232;  
RA van Harmelen J., Williamson C., Kim B., Morris L., Carr J.,  
RA Abdool Karim S.S., McCutchan F.;  
RT "Characterization of full-length HIV type 1 subtype C sequences from  
RT South Africa.";  
RL AIDS Res. Hum. Retroviruses 17:1527-1531(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA van Harmelen J.H., Williamson C., Kim B., Morris L., Carr J.,  
RA Abdool Karim S.S., McCutchan F.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY043176; AAL05334.1; -.  
DR HSSP; Q70622; IHVN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HC2NFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polyprotein; Zinc; Zinc-finger.  
SQ SEQUENCE 492 AA; 55075 MW; 11966A2F3D272ED0 CRC64;

Alignment Scores:

Pred. No.: 3.4e-100 Length: 492  
Score: 2437.00 Matches: 454  
Percent Similarity: 96.96% Conservative: 24  
Best Local Similarity: 92.09% Mismatches: 13  
Query Match: 87.91% Indels: 2  
DB: 2 Gaps: 2

US-09-475-704A-3 (1-1479) x Q901X0 (1-492)

QY 1 ATGGGGCGCGCGCGCAAGCATCTGCGCGCGCGCAAGCTGACCGCTGGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrrGluLysIleArg 20  
QY 61 CTGCGCGCGCGCGCAAGAGTGTATCATGATGAAGCACTGTGTGGCCAGCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysCysTyrMetLeuLysHisIleIleTrrAlaSerArgGlu 40

QY 121 CTGGAGAAGTTCCCTGAACCCCGGCTGTGGAGACCGAGGGCTGCAAGCAGATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrLysGluGlyCysLysGlnIle 60  
QY 181 ATCCGCCAGCTGCAACCCCGCTGCAGACCGCGCAGCGAGAGCTGAAGAGCTGTCAAC 240  
DB 61 IleAsnGlnLeuHisProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCTGTCACGAGAAGATCGAGTCCGCGACCAAGAGAGGCC 300  
DB 81 ThrValAlaThrLeuTyrCysValHisAlaGluIleGluValArgAspThrLysGluVal 100  
QY 301 CTGGACAAGATCGAGGAGGAGCAACAAGTCCAGCAGAAATCCAGCAGCGCCAGGCC 360  
DB 101 LeuAspLysIleGluGluGlnGlnAsnIleGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACTACCTCATCTGTGCAAACTGACGGCCAGATG 420  
DB 121 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 139  
QY 421 GTGCACCAAGGCATCAGCCCGCGCACCTGAAAGCTGTGTAAGGTGATCGAGAGAAG 480  
DB 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrrValLysValIleGluGluLys 159  
QY 481 GCCTTCAGCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 179  
QY 541 GACCTGAACACGATGTTGAACACCGTGGGGCGCCACGAGCGCCATGCGATGCTGAAG 600  
DB 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAACGAGAGCGCGCGAGTGGACCGCGTGACCCCGTGACCGCGGCC 660  
DB 200 AspThrIleAsnGluGluAlaAlaGluTrrAspArgLeuHisProValHisAlaGlyPro 219  
QY 661 ATGCCCCCGCGCAGATGCGCGGAGCCCGCGCGCAGCAGATCGCGCGCAGCAGCACC 720  
DB 220 ValAlaProGlyGlnMetArgLupProArgGlySerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGCAGAGCAGATCGCTGTGATGACCAACCCCCCATCCCGGTGGCGCATCTAC 780  
DB 240 LeuGlnGluGlnIleAlaTrrMetThrAlaAsnProProIleProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGTGGCTGAACAAGATCGTGCAGATGTACAGCCCGTGAGCATC 840  
DB 260 LysArgTrrIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 279  
QY 841 CTGCACATCAAGCAGGGCGCCCAAGAGCCCTCCGCACTACGTGACCGCTTCTTCAAG 900  
DB 280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTrrValAspArgPhePheLys 299  
QY 901 ACCGTGGCGCGCAGAGACCAACCGAGAGGTGAAGACTGATGACCGCAGACCCCTGCTG 960  
DB 300 ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrrMetThrGluThrLeuLeu 319  
QY 961 GTGCAGAACGCCCAACCCCGAAGTGAAGACCATCTGCGGCTCTCGGCCCGCGCCAGC 1020  
DB 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 339  
QY 1021 CTGGAGAAGATGATGACCGCTGCGCAGGGCGGTGGCGGCCCGCAGCAGCGCGCTG 1080  
DB 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyLysProGlyHisLysAlaArgVal 359  
QY 1081 CTGGCCGAGCGATGAGCCAGGCCAACACC--AGCGTATGATGACAGAGCAACTTC 1137  
DB 360 LeuAlaGluAlaMetSerGlnAlaAsnAsnIleAsnIleMetMetGlnLysSerAsnPhe 379  
QY 1138 AAGGGCCCGCGCGCATGTCAGTGTCTTCAACTGCGGCAAGAGGGCGCATGCGCCGC 1197  
DB 380 LysGlyProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGGCCACAGATG 1257

Db 400 AsnCyatrgAlaProatrgLysbGlyCybTrpLysCybGlyLysbGlnHbGlnMet 419  
QY 1258 AAGACTGCACCGAGCGCCAGCCAACTCTGGAGAATCTGGCCAGCCACAAGGCG 1317  
Db 420 LysAspCysIleGlnArgGlnAlaAsnHleuGlyLysIleTrpProSerHbLysGly 439  
QY 1318 CGCCCGGCAACTCTCTGACAGAGCCCGCCAGCCACCGCCCGCCCGAGAGCTTC 1377  
Db 440 ArgProGlyAsnPhelLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTCGAGAGACCAACCCCGCCAGACAGAGACAGACCGCCAGACCTGACC 1437  
Db 460 LysPheGlnGluThrThrProAlaProLysbGlnGluThrLysAspArgGluProLeuThr 479  
QY 1438 AGCCTGAAGAGCCTGTTCGCAACGACCCCTGAGCCAG 1476  
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

## RESULT 10

Q8AFW4 PRELIMINARY; PRT; 492 AA.  
ID Q8AFW4  
AC Q8AFW4;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22525961; PubMed=12639249; DOI=10.1089/088922203762688649;  
RA Williams C., Morris L., Maughan M.F., Ping L.-H., Dryga S.A.,  
RA Thomas R., Reap E.A., Chatters T., van Harmelen J., Pascual A.,  
RA Ramjee G., Gray G., Johnson R., Karim S.A., Swanson R.;  
RT "Characterization and selection of HIV-1 subtype C isolates for use in  
RT vaccine development."  
RL AIDS Res. Hum. Retroviruses 19:133-144(2003).  
DR EMBL; AF544010; AA016234.1; -.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR00721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 492 AA; 55221 MW; B1922020A1AE861B CRC64;

## Alignment Scores:

Pred. No.: 4.17e-100 Length: 492  
Score: 2435.00 Matches: 458  
Percent Similarity: 96.96% Conservative: 20  
Best Local Similarity: 92.90% Mismatches: 13  
Query Match: 87.84% Indels: 2  
DB: 2 Gaps: 2

US-09-475-704A-3 (1-1479) x Q8AFW4 (1-492)

QY 1 ATGGGCGCCCGCGCGCATCTCGCGCGCGGCAAGCTGAGCGCTGGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGluLysLeuAspLysTrpGluLysIleArg 20

QY 61 CTGGCCCCGCGGCAAGAGTGCTACATGATGAACACCTGTGTGGCCAGCCGCGAG 120  
Db 21 LeuArgProGlyGlyLysbHbTrpMetLeuLysbIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAGAAGTTCCCTGAAACCCCGGCTGTGGAGACCAAGAGGCTGCAAGAGATC 180  
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysbGlnIle 60  
QY 181 ATCCGACAGCTGACACCCCGCCCTGACAGCCGCGAGAGAGCTGAAGACTGTCAAC 240  
Db 61 MetLysbGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGluLeuLysSerLeuTrpAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGGTCCGCGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnLysIleGluValArgAspThrLysGluAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTGCCGACGAGAAGATCCAGAGCGCGAGGCC 360  
Db 101 LeuAspLysIleGlnGlnGlnAsnLysCysbGlnGlnLysThrGlnGlnAlaLysbAla 120

QY 361 GCCGCAAGAGGCGAAGTGAGCCAGAACTACCCCATCGTCAGAACTGACAGGCGCAGATG 420  
Db 121 AlaAsp--GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 139  
QY 421 GTGACCAAGGCGCATACGCCCCGACCTGAAACGCTGGGTGAAGTGATCGAGAGAG 480  
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 159

QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCCAG 540  
Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 179

QY 541 GACCTGAACACGATGTGAACACCGTGGCGGCGCCACCAAGCCCGCATCGAGTGTGAAG 600  
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199

QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGCAACCCCGTGACGCGCGCCC 660  
Db 200 AspThrIleAsnGln\*\*AlaAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 219

QY 661 ATCGCCCCGCGCAGATGCGCGAGCCCGCGCGCAGCATCGCGCGCACCAACGACACC 720  
Db 220 IleAlaProGlnGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239

QY 721 CTGCAAGAGCAGATGCGCTGTGATGACCAAGACCCCGCATCCCGGTGGCGACATCTAC 780  
Db 240 LeuGlnGlnGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyr 259

QY 781 AAGCGGTGATCATCTCTGGCGCTGAACAAGATCGTGGATGTACAGCCCGTGAGCATC 840  
Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIle 279

QY 841 CTGGAATCAAGCAGGCGCCCAAGAGAGCCCTTCGCGACTACGTGAGCCGCTTCTCAAG 900  
Db 280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299

QY 901 ACCGTGCGCGCGCAGACAGACCAAGAGGTGAAGAATGTGATGACCGACACCTGTCTG 960  
Db 300 ThrLeuArgAlaGlnGlnAlaThrGlnGlnValLysAsnTrpMetThrAspThrLeuLeu 319

QY 961 GTGCAAGACGCCAACCCGACTGCAAGACCATCTGCGCGCTTCGGCCCGCGCGCAGC 1020  
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339

QY 1021 CTGAGAGAGATGATGACCGCTGCGAGGCGCGTGGCGCGCCAGCAAGGCGCGCTG 1080  
Db 340 LeuGlnGlnMetMetThrAlaCysbGlnGlyValGlyGlyProGlyHisLysbAlaArgVal 359

QY 1081 CTGGCCGAGGCGATGAGCCAGGCAACACC--AGCGTATGATGACAGAGACCAACTTC 1137  
Db 360 LeuAlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPhe 379

QY 1138 AAGGCCCCCGGCATGCTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATGCCCCG 1197  
Db 380 LysGlyProArgArgIleValIlyScyPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
QY 1198 AACTGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCAGATG 1257  
Db 400 AsnCyArGAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
QY 1258 AAGGACTGCACCGAGCGCGCAAGGCCACTTCTGGGCAAGATCTGGCCCAAGCCACAGAGGC 1317  
Db 420 LysAspCysThrGluArgGlnAlaAsnPhenLeuGlyLysIleTrpProSerHisLysGly 439  
QY 1318 CGCCCCCGCACTCTCTGCAAGAGCGCGCGGAGCCCGGAGCCCGCGGAGAGCTTC 1377  
Db 440 ArgProGlyAsnPhenLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTCGAGAGAGACCCCGCGGCAAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
Db 460 ArgPheGluGluThrThrProAlaProLysGlnGluProIleGluArgGluProLeuThr 479  
QY 1438 AGCCTGAAGAGCGCTGTTCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476  
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 11

Q8UTH9 PRELIMINARY; PRT; 499 AA.  
ID Q8UTH9  
AC Q8UTH9;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP MEDLINE=21988475; PubMed=11991972;  
RX DOI=10.1128/JVI.76.11.5435-5451.2002;  
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,  
RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,  
RA Foley B.T., GaoLekwe S., Rybak N., Gaseitsiwe S., Vanberg F.,  
RA Marlink R., Lee T.-H., Essex M.;  
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:  
RT consensus sequence for an AIDS vaccine design?";  
RL J. Virol. 76:5435-5451(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,  
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vanberg F.,  
RA Marlink R., Lee T.-H., Essex M.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF443087; AAL34670.1; -.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; zf\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.  
SQ SEQUENCE 499 AA; 55793 MM; E573AF16EEBFB7F CRC64;

Alignment Scores:

Pred. No.: 5.37e-100 Length: 499  
Score: 2432.50 Matches: 455  
Percent Similarity: 95.99% Conservative: 24

Best Local Similarity: 91.18% Mismatches: 13  
Query Match: 87.75% Indels: 7  
DB: 2 Gaps: 2

US-09-475-704A-3 (1-1479) x Q8UTH9 (1-499)

QY 1 ATGGGCGCGCGCGGCAAGTCTGCGCGGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspThrTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGGCAAGAGTGTACATGATGAAGCACCTGGTGTGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisTyrMetIleLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGAGTTCGCGCTGAACCGCGCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATCCGCGAGCTGACACCGCGCTGACAGACCGGCGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCTGCTGACAGAGAGATCGAGGTTCGCGACACCAAGAGGCC 300  
Db 81 ThrIleAlaThrLeuTyrCysValHisGluSerIleGluValAlaArgAspThrLysGluAla 100  
QY 301 CTGAGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnAsnLysCysGlnGlnLysThrGlnGlnAlaGluAla 120  
QY 361 GCCGCAAGAGGGAGAGGTGAGCGCAAGACTACCCCATCGTGACAGAACTGACAGGGCCAGATG 420  
Db 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 140  
QY 421 GTGCACCAAGGCCATCAGCCCGCGCACCCCTGAACGCTGGGTGAAGGTGATGAGAGAGAG 480  
Db 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValAlaGluGluLys 160  
QY 481 GCCTTCAGCCCGGAGGTGATCCCATGTTACCGCCCTGAGCGAGGGCGGCGGCGGCGGAG 540  
Db 161 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 180  
QY 541 GACCTGAACAGATGTTGAACACCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
Db 181 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 200  
QY 601 GACACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 201 AspThrIleAsnGluGluAlaAlaGluTrpAspArgThrHisProValHisAlaGlyPro 220  
QY 661 ATGCGCGCGCGGCGAGATGCGGAGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720  
Db 221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240  
QY 721 CTGCAGGAGAGATGCGCTGTGATGACAGCAACCGCCCATCCCGTGGGCGGAGATCTAC 780  
Db 241 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProValProValGlyGluIleTyr 260  
QY 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGGATGACAGCCCGGTGAGCATC 840  
Db 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 280  
QY 841 CTGACATCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 281 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 300  
QY 901 ACCCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 301 ThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsnTrpMetThrAspThrLeuLeu 320  
QY 961 GTGCAGAGCGCAACCGCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340

QY 1021 CTGAGGAGATGATGACCGCCTGCGCAGGCGTGCGGCCAGCCCAAGGCCCGCTG 1080  
:::|||||  
Db 341 IIEGIUGIUMeTmEThrAlaCySGInGIyValGIyGlyProSerHISlySAlaArgVal 360  
QY 1081 CTGCGCGAGGCGATGAGCCAGGCCAAC--ACCAGCGTGATGATGCGAAGAGCAACTTC 1137  
|||||  
Db 361 SerAlaGIuAlaMeSerGIAlaAaNGIAlaAaNIleMeTmGIuLySserAaNPhe 380  
QY 1138 AAGGCGCCCGCGCGCATCGTCAAGTGTCTCAACTGCGCAAGAGGCCACATCGCCCGC 1197  
|||||  
Db 381 LySGIyProlYsArgIleValIyCySPheAaNCySGIyLySGIuGIyHISIleAlaArg 400  
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGCGTGCTGGAAGTGGCGCAAGAGGCCACAGATG 1257  
|||||  
Db 401 AaNCySArgAlaProlArgIyLySGIyCySTryLyCySGIyArgGIuGIyHISGIuMet 420  
QY 1258 AAGACTGCACCGAGCGCCAGGCCCACTTCCTGGCAAGATCTGGCCAGCCACAGAGGC 1317  
|||||  
Db 421 LySAspCySThrGIuArgGIAlaAaNPheLeuGIyLySIleTrProSerAaNIySGIy 440  
QY 1318 CGCGCGCGCAACTTCCTGCAGAGCGCGCGC-----GAGCCCGCGCC 1359  
|||||  
Db 441 ArgProGIyAaNPheLeuGIuAaNPArgProGIuProThrAlaProLeuGIuProThrAla 460  
QY 1360 CCGCGCGCGCGAGAGCTTCCTCGCTTCGAGAGAGACCAACCGCGCGCAAGAGAGCAAG 1419  
|||||  
Db 461 ProProlaGIuSerPheArgPheGIuGIuThrAaNProlaProLySGIuGIuLeuLyS 480  
QY 1420 GACCGCGAGACCCCTGACCAAGCCTGTTCGAGCAACGACCCCGCTGAGCCAG 1476  
:::|||||  
Db 481 AaNPArgIuThrLeuThrSerLeuArgSerLeuPheGIyGIyAaSProlLeuSerGIu 499

RESULT 12  
Q8UT71 PRELIMINARY; PRT; 492 AA.  
ID 08UT71;  
AC 08UT71;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21988475; PubMed=11991972;  
DOI=10.1128/JVI.76.11.5435-5451.2002;  
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,  
Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,  
Foley B.T., GaoLekwe S., Rybak N., Gaseitsiwe S., Vanberg F.,  
Marlink R., Lee T.-H., Essex M.;  
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:  
RT consensus sequence for an AIDS vaccine design?";  
RT J. Virol. 76:5435-5451 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,  
Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vanberg F.,  
Marlink R., Lee T.-H., Essex M.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF443099; AAL34778.1; -.  
DR HSSP; P24736; INCP.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf\_CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.

KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 492 AA; 55143 MW; FBFF5ABF180D9979 CRC64;  
Alignment Scores:  
Pred. No.: 5,66e-100 Length: 492  
Score: 2432.00 Matches: 455  
Percent Similarity: 96.75% Conservative: 22  
Best Local Similarity: 92.29% Mismatches: 14  
Query Match: 87.73% Indels: 2  
DB: 2 Gaps: 2  
US-09-475-704A-3 (1-1479) x Q8UT71 (1-492)

QY 1 ATGGCGCGCGCGCGCAGCATCTGCGCGCGCGCAAGCTGACCGCTGGAGCGCATCCGC 60  
|||||  
Db 1 MetGIyAlaArgAlaSerValLeuLySGIyGIuLySLeuAaPrlaTrpAlaSerArgIu 20  
QY 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGGTGTGGCGCGCGAG 120  
|||||  
Db 21 LeuArgProGIyGIuLyLySGIyTyMetLeuLySHISleuValTrpAlaSerArgIu 40  
QY 121 CTGGAAGAAGTTCGCGCTGAACCGCGCGCTGCTGAGACCGAGCGAGAGCTGAAGCCTGTTCAAC 180  
|||||  
Db 41 LeuGIuArgPheAlaLeuAaNProlYleuLeuGIuThrSerGIuGIyCySArgGIuIle 60  
QY 181 ATCGCGCAGCTGCACCGCGCGCTGCAGACCGCGCGAGAGAGCTGAAGCCTGTTCAAC 240  
:::|||||  
Db 61 MetArgGIuLeuGIuProAlaLeuGIuThrGIyThrGIuGIuLeuArgSerLeuPheAa 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAGATCGAGTCCGCGCACCAAGAGGCC 300  
|||||  
Db 81 ThrValAlaThrLeuTyrcySValHISGIuLySIleProValArgAaSPThrLySGIuAla 100  
QY 301 CTGGAACAAGATCGAGAGAGAGAGCAACAAGTCCAGCAAGATCCAGAGCGCGAGGCC 360  
|||||  
Db 101 LeuAaSPlySIleGIuGIuGIuAaNPlyCySGInGIuLyThrGIuGIuThrGIuAla 120  
QY 361 GCCGACAAGGCGAAGTGAAGCCCAACTACCCCATCTGCAGAACCTGCAGAGCGCGAGATG 420  
|||||  
Db 121 Ala--AlaGIyLySValSerGIuAaNPtyrProIleValGIuAaNPleuGIuGIuMet 139  
QY 421 GTGCAACGAGCCATCAGCGCGCGCAACCTGTAAGCCTGGTGAAGTGATCGAGAGAAG 480  
|||||  
Db 140 ValHISGIuAlaIleSerProArgThrLeuAaNPAlaTrpValLySValIIEGIuGIuLyS 159  
QY 481 GCCTTCAGCGCGCGAGGTGATCCCATGTTCACCGCGCTGAGCGAGCGCGCCAGCCCCAG 540  
|||||  
Db 160 AlaPheSerProGIuValIleProMetPheThrAlaLeuSerGIuGIyAlaThrProGIu 179  
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGCGCAACGCGCGCATGCAAGTGTGAAG 600  
|||||  
Db 180 AaSPLeuAaNPThrMetLeuAaNPThrValGIyGIyHISGIuAlaIleMetGIuMetLeuLyS 199  
QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCGCGGTGCACCGCGGTGCACGCGCGCCCC 660  
|||||  
Db 200 AaSPThrIleAaNPGIuGIuAlaIleGIuTrpAaPArgValHISProValHISAlaGIyPro 219  
QY 661 ATGCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGAGCATGCGCGCGCACCGAGCACACC 720  
|||||  
Db 220 IleAlaProGIyGIuMetArgIuProArgIySerAaPrlaIleAGIyThrThrSerThr 239  
QY 721 CTGCAAGAGCAGATCGCCTGTGATGACCAACCGCGCATCCCGGTGGCGAGCATCTAC 780  
|||||  
Db 240 LeuGIuGIuGIuIleAlaIleAlaTrpMetThrGIyAaNPProValProValGIyAaSPlyeTy 259  
QY 781 AAGCGGTGATCATCTGCGCGCTGAACAAGATGTGGATGTACAGCCCGGTGAGCATC 840  
|||||  
Db 260 LySArgTrpIleIleLeuGIyLeuAaNPlySIleValArgMetTySerProValSerIle 279  
QY 841 CTGGAACATCAAGAGAGCGCGCAAGAGCGCTTCGCGAGTACGTGAGCGCTTCTTCAAG 900  
|||||  
Db 280 LeuAaSPlyeArgGIuGIyProLySGIuProPheArgAaSPtyrValAaSPArgPheNHelyS 299

```
QY 901 ACCCTGCGCGCCGAGCAGACCCGAGAGGTGAAGACTGATGACCGACACCTGCTG 960
    |||||||
Db 300 ThrLeuArgAlaGlnGlnAlaThrGlnAsnProValLysAsnTrpMetThrAspThrLeu 319
QY 961 GTGCAGAACGCCAACCCGACTGCAAGACCATCTCGCGCTCTCGCCCCGCGCCAGC 1020
    |||||||
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339
QY 1021 CTGAGAGAGATGATGACCGCTGCGCAGGCGGTGGCGGCCCGACCAAGGCCGCTG 1080
    |||||||
Db 340 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359
QY 1081 CTGCGCGAGCGCATGAGCCAGGCC--AACACCAAGCTGATGATGCAAGACCACTTC 1137
    |||||||
Db 360 LeuAlaGlnAlaMetSerGlnAlaAsnAlaAsnIleMetMetGlnArgSerAsnPhe 379
QY 1138 AAGGCGCCCGCGCGCATCTGCAAGTGTCTCAACTGCGGCAAGAGGCGCCACATCGCCCGC 1197
    |||||||
Db 380 LysGlyProArgLysIleIleLysCysPheAsnCysGlyLysGlnGlyHisLeuAlaArg 399
QY 1198 AACTGCGCGCGCGCCCGCAGAGAGGCTGTGAACTGCGGCAAGAGGCGCCACGATG 1257
    |||||||
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisMet 419
QY 1258 AAGGACTGACCGCAGCGCCAGGCAACTTCTGCGCAAGATCTGGCCAGCCAGAGGCG 1317
    |||||||
Db 420 LysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439
QY 1318 CGCCCCGCGCAACTTCTGCAAGAGCGCGCCCGCAGCCCGCCCGCGCGAGACTTC 1377
    |||||||
Db 440 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGlnSerPhe 459
QY 1378 CGCTTCAGAGAGACCAACCCCGCCAGAGAGAGCAAGACCGCGAGACCTGACC 1437
    |||||||
Db 460 ArgPheGlnGluThrThrProAlaProLysGlnGlnLysAspArgGluProLeuThr 479
QY 1438 AGCCTGAAGAGCCTGTTCGCAACGACCCCTGAGCCAG 1476
    |||||||
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492
```

## RESULT 13

```
Q8UT53
ID 08UT53 PRELIMINARY; PRT; 492 AA.
AC 08UT53;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Gag protein.
GN Name=gag;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21988475; PubMed=11991972;
RX DOI=10.1128/JVI.76.11.5435-5451.2002;
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,
RA Williams C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,
RA Foley B.T., Gaolekwe S., Rybak N., Gaseitsiwe S., Vanberg F.,
RA Marlink R., Lee T.-H., Essex M.;
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:
RT consensus sequence for an AIDS vaccine design?";
RL J. Virol. 76:5435-5451(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vanberg F.,
RA Marlink R., Lee T.-H., Essex M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF443101; AAL34796.1; -.
DR HSSP; Q70622; 1HVN.
DR GO; GO:0019012; C:viriion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
```

```
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; ZNF_C2HC; 2.
DR PROSITE; PS50158; ZF_CCHC; 2.
KW AIDS; Core protein; Polyprotein.
SQ SEQUENCE 492 AA; 55087 MW; 9B80F1A1426E80C9 CRC64;

Alignment Scores:
Pred. No.: 7,68e-100 Length: 492
Score: 2429.00 Matches: 457
Percent Similarity: 96.75% Conservative: 20
Best Local Similarity: 92.70% Mismatches: 14
Query Match: 87.63% Indels: 2
DB: 2 Gaps: 2

US-09-475-704A-3 (1-1479) x Q8UT53 (1-492)
QY 1 ATGGGCGCGCGCCAGCATCTGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60
    |||||||
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspThrTrpGluLysIleArg 20
QY 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGACACTGTGTGGCCAGCCGCGAG 120
    |||||||
Db 21 LeuArgProGlyGlyLysLysArgTyrMetLeuLysHisLeuIleTrpThrSerArgGlu 40
QY 121 CTGAGAGAAGTTGCGCCTGAACCCCGCGCTGTGAGAGACCGAGCGGTGCAAGCATC 180
    |||||||
Db 41 LeuGluArgPheAlaLeuAsnProAspLeuLeuGluThrSerGlnGlyCysLysGlnIle 60
QY 181 ATCCGCCAGCTGCACCCCGCGCTGCAGACCGCGCAGAGAGCTGAAGAGCTGTTCAC 240
    |||||||
Db 61 IleGlnGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAGATCGAGGTCCGACACCAAGAGGCC 300
    |||||||
Db 81 ThrValAlaThrLeuTyrCysValHisLysGlyIleGluValGlnAspThrLysGluAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTGCAGACAGATCCAGACGCGGAGGCC 360
    |||||||
Db 101 LeuAspLysIleGlnGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaGluAla 120
QY 361 GCCGACAAGGCGAAGGTGAGCCAGAACTACCCCATCGTGCAAACTGACGCGCCAGATG 420
    |||||||
Db 121 Ala--AlaGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 139
QY 421 GTGCACCAAGGCCATCAGCCCGCGCACCTGAAAGCCTGGTGAAGGTGATCGAGAGAAG 480
    |||||||
Db 140 ValHisGlnSerIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGluLys 159
QY 481 GCCTTCAGCCCGAGGTGATCCCATGTTGACCGCCCTGAGCGAGGCGCCACCCCCAG 540
    |||||||
Db 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 179
QY 541 GACCTGAACAGATGTTGAACACCGGTGGCGGCCACCAAGCGCCCATGACATGCTGAAG 600
    |||||||
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyHisGlnAlaAlaMetGlnMetLeuLys 199
QY 601 GACACCATCAAGAGAGGCGCGCGAGTGGACCGCGTGCACCCCGTGACGCGCGGCC 660
    |||||||
Db 200 AspThrIleAsnGlnGlnAlaGlnAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 219
QY 661 ATCGCCCCCGCGAGATGCGCGAGCCCGCGCGAGCGACATCGCCGCGACCAAGCACC 720
    |||||||
Db 220 IleAlaProGlyGlnLeuArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239
QY 721 CTGCAGAGAGATGCGCTGATGACCAAGAACCCCCCATCCCGGTGGGCGACATCTAC 780
    |||||||
Db 240 LeuGlnGlnGlnIleAlaTrpMetThrAsnAsnProProValProValGlyAspIleTyr 259
QY 781 AAGCGGTGATCATCTCGGCGCTGAACAAGATCGTGGAGATGTACAGCCCGTGAGCATC 840
```

|||||  
Db 260 LysArgTRpIleIleLeuGlyLeuAenlySileValArgMetTySerProValSerIle 279  
QY 841 CTGCATCATCAGCAGGAGCCCAAGAGCCCTTCCGGCACTAGCTGACCGCTTCTTCAAG 900  
|||||  
Db 280 LeuAepIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCCGAGCAGACCCCAAGAGGTGAAGAACTGGATGACCGACACCTGCTG 960  
|||||  
Db 300 ThrLeuArgAlaGlnGlnSerSerGlnGluValLysAsnTrpMetThrAspThrLeu 319  
QY 961 GTGCAGAACGCCAACCCCGACTGCAGAACCATCTCGCGCTCTCGGCCCCGGCGCAGC 1020  
|||||  
Db 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339  
QY 1021 CTGGAGAGATGATGATACCGCTTGCAGGGCGCTGGGCGGCCCAAGAGCCCGCGTG 1080  
|||||  
Db 340 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgAla 359  
QY 1081 CTGGCCGAGGCGGATGATGAGCCAGGCC--AACAACAGCGTGATGATGCAGAAAGCAACTTC 1137  
|||||  
Db 360 LeuAlaGlnAlaMetSerGlnAlaAsnAsnThrAsnIleMetMetGlnLysSerAsnPhe 379  
QY 1138 AAGGGCCCCCGGCGCATGCTCAAGTGTTCAACTGCGGCAAGAGGGCCACATCGCCCGC 1197  
|||||  
Db 380 LysGlyProArgArgIleValLysCysPheAsnCyGlyLysGlnGlyHisIleAlaArg 399  
QY 1198 AACTGCGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTGGCGCAAGAGGGCCACAGATG 1257  
|||||  
Db 400 AsnCySArgAlaProArgLysLysGlyCysTrpLysCyGlyLysGlnGlyHisIleMet 419  
QY 1258 AAGGACTGCACCGGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCAGCCACAAAGGGC 1317  
|||||  
Db 420 LysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439  
QY 1318 CGCCCCGGCAACTTCTGCAGAGCGCGCCCGAGCCACCGCCCGCGCGAGAGCTTC 1377  
|||||  
Db 440 ArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTCGAGAGAGACCAACCCCGCGCAAGAGCAGAGCAAGCAACCGCAGACCTGACC 1437  
|||||  
Db 460 ArgPheGlnGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuThr 479  
QY 1438 AGCCTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAG 1476  
|||||  
Db 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492

RESULT 14

Q7SLX7 PRELIMINARY; PRT; 491 AA.  
AC Q7SLX7;  
DT 01-OCT-2003 (TReMBLrel. 25, Created)  
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gupta R.M., Prasad V., Singh B., Seth P.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF531311; AAP87971.1; -.  
DR HSSP; Q70622; IHVN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retroviral matrix.  
DR InterPro; IPR000071; Retroviral p17.  
DR InterPro; IPR008916; Retroviral capsid\_C.

DR InterPro; IPR008919; Retroviral capsid\_N.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; Zf\_CCHC; 2.  
KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 491 AA; 54815 MW; 36CD9A56CAEPAD7C CRC64;

Alignment Scores:

Pred. No.:	8.94e-100	Length:	491
Score:	2427.50	Matches:	455
Percent Similarity:	96.75%	Conservative:	21
Best Local Similarity:	92.48%	Mismatches:	15
Query Match:	87.57%	Indels:	1
DB:	2	Gaps:	1

US-09-475-704A-3 (1-1479) x Q7SLX7 (1-491)

QY 1 ATGGAGCGCGCGCCAGCATCTCGCGCGCGGCAAGCTGACGCTGGAGCGCATCCGC 60  
|||||  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCCCGCGCGCAAGAAGTGTACATGATGAACACCTGTGTGGCCAGCGCGAG 120  
|||||  
Db 21 LeuArgProGlyGlyLysLysHisLysTrpMetLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGGAGAAGTTGCGCCTGAACCCCGCGCTGTGAGAGCAGCGAGGCTGCAAGCAGATC 180  
|||||  
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysLysGlnIle 60  
QY 181 ATCGCCAGCTGACACCCCGCTGCAGACCGGCGAGAGAGCTGAAGCTGTTCAC 240  
|||||  
Db 61 LeuLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300  
|||||  
Db 81 ThrValAlaThrLeuTyrCysValHisAlaGlyTleGluValArgAspThrLysGluAla 100  
QY 301 CTGCACAGATCGAGAGAGAGCAGAACAAAGTCCAGCAGAGATCCAGAGCGCGAGCC 360  
|||||  
Db 101 LeuAspLysIleGlnGlnGlnGlnAsnLysGlyGlnGlnLysThrGlnGlnAlaLysGly 120  
QY 361 GCCGACAAAGGCGAAGGTGAGCCAGAACTACCCCATCGTGCAAGAACTGCAGGGCCAGATG 420  
|||||  
Db 121 AlaAsp--GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 139  
QY 421 GTGCACCAAGGCCATCAGCCCCCGCAACCTGAACGCTGGGTGAAGTGATCGAGAGAG 480  
|||||  
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGluLys 159  
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
|||||  
Db 160 AlaPheSerProGluValIlePrometPheThrAlaLeuSerGlnGlyAlaThrProGln 179  
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGGCCCAAGGCCCGCATGCAGATGCTGAAG 600  
|||||  
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLys 199  
QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCTGCACCCCGTGCACGCGCGCCC 660  
|||||  
Db 200 AspThrIleAsnGlnGluAlaIleGluTrpAspArgLeuHisProValHisAlaGlyPro 219  
QY 661 ATGGCCCCCGGCGAGATGGCGGAGCCCGCGGAGCGAGACATCGCGCGCACCAACGAGACC 720  
|||||  
Db 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGCAGAGCAGATCGCCTGTGATGACCAAGCAACCCCCCATCCCGGTGGCGACATCTAC 780  
|||||  
Db 240 LeuGlnGlnGlnIleAlaTrpMetThrAsnAsnProProValProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTCGGCGCTGAACAAGATCGTGGAGATGTACAGCCCGTGACATC 840

```
|||||
Db      260 LysArgTrpIleIleuGlyLeuAnlysIleValArgMetTyrSerProValSerIle 279
QY      841 CTGACATCAAGCAGGGCCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAG 900
Db      280 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299
QY      901 ACCCTGGCGCCGAGACAGACCCCAAGAGGTGAAGAACTGGATGACCGACACCTGCTG 960
Db      300 ThrLeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319
QY      961 GTGCAGAACGCCAACCCGACTGCAAGACCATCTCGCGCTCTCGGCCCGCGCCAGC 1020
Db      320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339
QY      1021 CTGAGAGAGATGATGACCGCCTGCGCAGGGCGTGGCGGCCCAAGGCCGCGTG 1080
Db      340 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 359
QY      1081 CTGGCCGAGGCGATGAGCCAGCCACAACACCGCTGATGATGACAGAGAACATTCAG 1140
Db      360 LeuAlaGlnAlaMetSerGlnThrGlySerThrIleMetMetGlnArgSerAsnPheLys 379
QY      1141 GGGCCCCGGCGCATGTCAGTGTCAACTGCTCACTGCGGCAAGAGGGCCACATCGCCGCAAC 1200
Db      380 GlySerLysArgIleValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaArgAsn 399
QY      1201 TGCCGCGCCCCCGCAAGAGGGCTGTGAGTGTGCAAGTCCGCAAGAGGCCACCATGAAG 1260
Db      400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisGlnMetLys 419
QY      1261 GACTGCACCGCAGCGCCAGGCCCACTTCTTGGGCAGATCTGGCCCAAGGCCACAGGGCGCG 1320
Db      420 AspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439
QY      1321 CCGCGCACTTCTCTGAGAGCGCGCCCGGAGCCCAAGGCCCGCCCGGAGAGCTTCCGC 1380
Db      440 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 459
QY      1381 TTCGAGGAGACCAACCCCGCGCAAGACGAGAGCAAGAGACCGCGCAACCTGACCAAGC 1440
Db      460 PheGlnGluThrThrProAlaProLysGlnGlnLeuLysAspArgGluProLeuThrSer 479
QY      1441 CTGAAGAGCCTGTTGGGCAACGACCCCTGAGCCAG 1476
Db      480 LeuLysSerLeuPheGlySerAspProLeuSerGln 491

RESULT 15
Q7SLX3 PRELIMINARY; PRT; 492 AA.
ID Q7SLX3;
AC Q7SLX3;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Gag protein.
GN Name=gag;
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Gupta R.M., Prasad V., Singh B., Seth P.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF53135; AAP87975.1; -.
DR HSSP; Q70622; 1HVN.
DR GO; GO:0019012; C:viriion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR010999; Retrovir_matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.
```

```
DR      InterPro; IPR008919; Retrov_capsid_N.
DR      InterPro; IPR001878; Znf_CCHC.
DR      Pfam; PF00540; Gag_p17; 1.
DR      Pfam; PF00607; Gag_p24; 1.
DR      Pfam; PF00098; ZF_CCHC; 2.
DR      SMART; SM00343; Znf_C2HC; 2.
DR      PROSITE; PS50158; ZF_CCHC; 2.
DR      AIDS; Core protein; PolypoteIn.
SQ      SEQUENCE 492 AA; 55105 MW; 78697BC7B49D3CD3 CRC64;

Alignment Scores:
Pred. No.:      1,28e-99      Length:      492
Score:          2424.00      Matches:      454
Percent Similarity: 95.93%      Conservative: 18
Best Local Similarity: 92.28%      Mismatches: 20
Query Match:    87.45%      Indels:      0
DB:              2      Gaps:      0

US-09-475-704A-3 (1-1479) x Q7SLX3 (1-492)
QY      1 ATGGGGCGCCGCGCCAGCATCTCGCGCGGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60
Db      1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGlnLysIleArg 20
QY      61 CTGCGCCCCGGCGCAAGAGTGTACATGATGAAGCACCTGTGTGGCCAGCGCGAG 120
Db      21 LeuArgProGlyGlyLysLysArgTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40
QY      121 CTGAGAAAGTTGCGCCCTGAACCCCGCGCTGTGAGAGACGAGGGCTGCAAGCAGATC 180
Db      41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerThrGlyCysLysGlnIle 60
QY      181 ATCCGCGAGCTGACACCCCGCTGACAGACCGGCGAGGAGAGCTGAAGAGCTGTCAAC 240
Db      61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuArgSerLeuPheAsn 80
QY      241 ACCGTGGCCACCTGTACTGCTGTGACGAGAGATCGAGTCCGCGACACCAAGAGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHisGlnGlyIleGlnValaArgAspThrLysGlnAla 100
QY      301 CTGACAAGATGAGAGAGAGAGCAACAAGTCCAGAGATCCAGAGCGCGAGGCC 360
Db      101 LeuAspLysIleGlnGlnGlnGlnAsnLysIleGlnGlnLysThrGlnGlnGlnLysGlu 120
QY      361 GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCGTGACAACTGACAGGGCGAGATG 420
Db      121 AlaAspGlyAsnGlnValaSerGlnAsnTyrProIleValaGlnAsnLeuGlnGlnMet 140
QY      421 GTGCACGAGGCCATCAGCCCGCGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAG 480
Db      141 ValHisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 160
QY      481 GCCTTCAGCCCGGAGGTGATCCCATGTTACCGCCCTGAGCGAGGGCGCCACCCCGCAG 540
Db      161 AlaPheSerProGlnValaIlePrometPheThrAlaLeuSerGlnGlyAlaThrProGln 180
QY      541 GACCTGAACACGATGTTGAACACCGTGGCGCGGCAACGCGCCATGACAGATGCTGAAG 600
Db      181 AspLeuAsnThrMetLeuAsnThrValaGlyGlyHisGlnAlaIleMetGlnMetLeuLys 200
QY      601 GACACCATCAAGAGAGAGCGCGCGAGTGGACCGCGTGACCCCGTGACCGCGGCCCC 660
Db      201 AspThrIleAsnGlnGlnAlaIleGlnAlaGlnTrpAspArgLysHisProValHisIleGlyPro 220
QY      661 ATCGCCCCCGCGCAGATGCGCGAGCCCGCGGCGAGCAGACATGCGCGGACCAAGCAGCC 720
Db      221 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 240
QY      721 CTGCAGAGAGATGCGCTGTGAAGACGAGAACCCCGCATCCCGGTGGCGCATCTAC 780
Db      241 LeuGlnGlnGlnIleAlaTrpMetThrSerAsnProProValProValGlyAspIleTyr 260
QY      781 AAGCGGTGATCATCTCTGGGCTGAACAAGATGCTGCGGATGTACAGCCCGTGAGCATC 840
```

```
Db      261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProAlaSerIle 280
QY      841 CTGCATCAAGCAGGGCCCCAAGGAGCCCTTCGGCACTACGTGACCGCTTCTTCAAG 900
Db      281 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 300
QY      901 ACCCTGCGCGCGAGCAGACACCAGAGGTGAAGACTGATGACCGACACCCCTGCTG 960
Db      301 ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 320
QY      961 GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCGCGCGCAGC 1020
Db      321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340
QY      1021 CTGAGGAGATGATGACCGCTGCGCAGGGCGTGCGCGGCCCAAGCGCGCGTG 1080
Db      341 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 360
QY      1081 CTGCGCCGAGGCGATGAGCCAGCCAAACACCAAGCGTGATGATGACAGAACCACTTCAAG 1140
Db      361 LeuAlaGluAlaMetSerGlnThrGlySerThrIleMetMetGlnArgSerAsnPheLys 380
QY      1141 GGGCCCCGCGCATGCTCAAGTGTCTCACTGCGGCAAGAGGGCCACATCGCCGCAAC 1200
Db      381 GlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 400
QY      1201 TGCCGCGCGCCCCCGCAAGAGGGCTGTGAAGTGGCCAAAGAGGGCCACCAAGATGAAG 1260
Db      401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420
QY      1261 GACTGCACCGGAGCGGCCCACTTCTGGGCAAGATCTGGCCCAAGCCACAGGGCGCG 1320
Db      421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY      1321 CCGGCAACTTCTGCAAGAGCGCGCCCGAGCCCAACCGCCCCCGCGGAGAGCTTCGGC 1380
Db      441 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 460
QY      1381 TTCGAGGAGACCAACCCCGCGCAAGACAGAGCAAGAGACCGCGGAGACCTGACCAAGC 1440
Db      461 PheGluGluThrThrProAlaProLysGlnGluLeuLysAspArgGluProLeuThrSer 480
QY      1441 CTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAG 1476
Db      481 LeuLysSerLeuPheGlySerAspProLeuSerGln 492
```

Search completed: March 11, 2005, 15:39:19  
Job time : 285.42 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 13:16:01 ; Search time 6600.56 Seconds  
(without alignments)  
10857.457 Million cell updates/sec

Title: US-09-475-704A-3  
Perfect score: 1479  
Sequence: 1 atggcgccgcgcgcgcacat.....acgacccctgagccagtaa 1479

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1479	100.0	1479	6	AX455887 Sequence
2	1479	100.0	1479	6	AX468543 Sequence
3	1473.4	99.6	1479	6	AX468547 Sequence
4	1463	98.9	1479	6	AX455904 Sequence
5	1328.2	89.8	1485	12	AY181195 Synthetic
6	1288.8	87.1	1491	6	AX455983 Sequence
7	1288	87.1	1494	6	AX455935 Sequence
8	1276.8	86.3	1509	6	AX455888 Sequence
9	1275.2	86.2	1509	6	AX468544 Sequence
10	1273.6	86.1	1509	6	AX468548 Sequence
11	1260.8	85.2	1509	6	AX455905 Sequence
12	1221.2	82.6	9166	6	AX427930 Sequence
13	1206	81.5	9788	6	AX427936 Sequence
14	1205.4	81.5	4288	6	AX149648 Sequence
15	1197.8	81.0	1509	12	AF201927 Synthetic
16	1197.8	81.0	1515	6	BD263637 Improved
17	1197.8	81.0	1515	6	CQ870496 Sequence
18	1197.8	81.0	1515	6	AR373320 Sequence
19	1197.8	81.0	4472	6	BD263697 Improved

20	1197.8	81.0	4472	6	CQ870567 Sequence
21	1197.8	81.0	4472	6	AR373380 Sequence
22	1197.8	81.0	4608	6	BD263698 Improved
23	1197.8	81.0	4608	6	CQ870568 Sequence
24	1197.8	81.0	4608	6	AR373381 Sequence
25	1197.8	81.0	4689	6	BD263696 Improved
26	1197.8	81.0	4689	6	CQ870566 Sequence
27	1197.8	81.0	4766	6	BD263695 Improved
28	1197.8	81.0	4766	6	CQ870565 Sequence
29	1197.8	81.0	4766	6	AR373378 Sequence
30	1197.8	81.0	4766	6	BD263640 Improved
31	1195.8	80.9	2031	6	CQ870499 Sequence
32	1195.8	80.9	2031	6	AR373323 Sequence
33	1195.8	80.9	2031	6	AY531263 Synthetic
34	1194.2	80.7	1503	12	AF287354 Synthetic
35	1159.6	78.4	1539	6	AX306431 Sequence
36	1134.4	76.7	1539	6	AX427919 Sequence
37	1131.8	76.5	6438	6	AX427925 Sequence
38	1131.8	76.5	9194	6	AX427926 Sequence
39	1131.8	76.5	9194	6	AX427927 Sequence
40	1131.8	76.5	12411	6	AX427928 Sequence
41	1120.4	75.8	9169	6	AX427921 Sequence
42	1116.6	75.5	9189	6	AX427932 Sequence
43	1116.6	75.5	9792	6	AX427932 Sequence
44	1116.6	75.5	9792	6	AF202464 Synthetic
45	1114.6	75.4	1847	12	AF202464 Synthetic

ALIGNMENTS

RESULT 1	AX455887	1479 bp	DNA	linear	PAT 06-JUL-2002
LOCUS	Sequence 3 from Patent WO0204493.				
DEFINITION	AX455887				
ACCESSION	AX455887.1	GI:21714880			
VERSION					
KEYWORDS					
SOURCE	synthetic construct				
ORGANISM	synthetic construct				
REFERENCE	1 zur Megede,J., Barnett,S.W., Engelbrecht,S., and van Rensburg,B.				
AUTHORS	Polynucleotides encoding antigenic hiv type c polypeptides,				
TITLE	Patent: WO 0204493-A 3 17-JAN-2002;				
JOURNAL	CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)				
FEATURES	Location/Qualifiers				
source	1.1479				
	/organism="synthetic construct"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:32630"				
	/note="synthetic Gag of HIV strain AF110965"				
ORIGIN					
Query Match	100.0%;	Score 1479;	DB 6;	Length 1479;	
Best Local Similarity	100.0%;	Pred. No. 9.5e-151;			
Matches 1479;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	ATGGGCGCCCGCGCCAGCATCTGCGCGCGGCAAGCTGAGCGCCTGGAGCGCATCCGC	60		
Db	1	ATGGGCGCCCGCGCCAGCATCTGCGCGCGGCAAGCTGAGCGCCTGGAGCGCATCCGC	60		
QY	61	CTGCGCCCGCGCGGCAAGAGTCTACATGATGAAGCACCTGTGTGGCCAGCGCGAG	120		
Db	61	CTGCGCCCGCGCGGCAAGAGTCTACATGATGAAGCACCTGTGTGGCCAGCGCGAG	120		
QY	121	CTGAGAGATTGCGCCCTGAACCCCGGCTGCTGAGAGACGAGGGCTGCAACAGATC	180		
Db	121	CTGAGAGATTGCGCCCTGAACCCCGGCTGCTGAGAGACGAGGGCTGCAACAGATC	180		
QY	181	ATCCGCGAGCTGCAACCCCGCTGAGAGCCGCGAGGAGGAGCTGAAGAGCTGTTCAC	240		

Db 181 ATCCGCCAGCTGACACCCCGCTGCAACCGGACGGAGAGCTGAAGACCTGTTCAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCTGTCAGAGAAGATGAGTCCGCGACACCAAGAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTACTGCTGTCAGAGAAGATGAGTCCGCGACACCAAGAGAGGCC 300  
QY 301 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 301 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAGACTACCCCATGCTGAGAACTGACAGGGCCAGATG 420  
Db 361 GCCGACAAGGGCAAGGTGAGCCAGAGACTACCCCATGCTGAGAACTGACAGGGCCAGATG 420  
QY 421 GTGCAACGAGCCATCAACCCCGCAACCTGAAACGCTGGGTGAAGGTGATCGAGAGAGAG 480  
Db 421 GTGCAACGAGCCATCAACCCCGCAACCTGAAACGCTGGGTGAAGGTGATCGAGAGAGAG 480  
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
Db 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCCACAGGGCCGCAATGAGATGCTGAAG 600  
Db 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCCACAGGGCCGCAATGAGATGCTGAAG 600  
QY 601 GACACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 601 GACACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 661 ATGCGCCCCGCGCAGATGCGCGAGCGCCGCGGAGCGAGCATCGCCGGCAACCAAGCAACC 720  
Db 661 ATGCGCCCCGCGCAGATGCGCGAGCGCCGCGGAGCGAGCATCGCCGGCAACCAAGCAACC 720  
QY 721 CTGCAAGAGAGAGATGCGCTGATGAGCAAGCAACCCCATCCCGTGGGCGAGCATCTTAC 780  
Db 721 CTGCAAGAGAGAGATGCGCTGATGAGCAAGCAACCCCATCCCGTGGGCGAGCATCTTAC 780  
QY 781 AAGCGGTGATCATCTCTGGGCTGGAACAAGATCGTGGATGTAACAGCCCGTGAAGCATC 840  
Db 781 AAGCGGTGATCATCTCTGGGCTGGAACAAGATCGTGGATGTAACAGCCCGTGAAGCATC 840  
QY 841 CTGGAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 841 CTGGAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 ACCCTGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 901 ACCCTGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
QY 1021 CTGGAAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
Db 1021 CTGGAAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1081 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
Db 1081 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 GGGCCCCGCGCATCTCAAGTGTCTCAACTGCGGCAAGAGAGGGCCACATCGCCGCAAC 1200  
Db 1141 GGGCCCCGCGCATCTCAAGTGTCTCAACTGCGGCAAGAGAGGGCCACATCGCCGCAAC 1200  
QY 1201 TGCCGCGCGCGCGCAAGAGGGGTGCTGAAGTGGGCAAGAGAGGGCCACCAAGTGAAG 1260  
Db 1201 TGCCGCGCGCGCGCGCAAGAGGGGTGCTGAAGTGGGCAAGAGAGGGCCACCAAGTGAAG 1260  
QY 1261 GACTGCAACGAGCGCAAGGCAACTTCTGGGCAAGATCTGGCCAGCCACCAAGGGCGCG 1320  
Db 1261 GACTGCAACGAGCGCGCAAGGCAACTTCTGGGCAAGATCTGGCCAGCCACCAAGGGCGCG 1320

QY 1321 CCGGCAACTTCTCTGACAGAGCGCGCCCGAGCCACCGCCCCCGCGAGAGCTTCCGC 1380  
Db 1321 CCGGCAACTTCTCTGACAGAGCGCGCCCGAGCCACCGCCCCCGCGAGAGCTTCCGC 1380  
QY 1381 TTGAGAGAGACACCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db 1381 TTGAGAGAGAGACACCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 CTGAAGAGCGCTGTCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
Db 1441 CTGAAGAGCGCTGTCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479

RESULT 2  
AX468543 1479 bp DNA linear PAT 16-JUL-2002  
LOCUS Sequence 63 from Patent WO0226209.  
DEFINITION AX468543  
ACCESSION AX468543  
VERSION AX468543.1 GI:21901373  
KEYWORDS  
SOURCE Human immunodeficiency virus 1 (HIV-1)  
ORGANISM Human immunodeficiency virus 1  
REFERENCE 1 O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
AUTHORS Singh,M., Uimer,J. and Dubensky,T.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 63 04-APR-2002;  
FEATURES  
source location/Qualifiers  
1. 1479  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"

Query Match 100.0%; Score 1479; DB 6; Length 1479;  
Best Local Similarity 100.0%; Pred. No. 9.5e-151;  
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
Db 1 ATGGGCGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
QY 61 CTGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
Db 61 CTGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
QY 121 CTGGAAGAGTTCGCTGAAACCGCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
Db 121 CTGGAAGAGTTCGCTGAAACCGCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
QY 181 ATCCGCAAGTGCACCCCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 181 ATCCGCAAGTGCACCCCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
QY 241 ACCGTGGCAACCTGTACTGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
Db 241 ACCGTGGCAACCTGTACTGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
QY 301 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 301 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAGACTACCCCATGCTGAGAACTGACAGGGCCAGATG 420  
Db 361 GCCGACAAGGGCAAGGTGAGCCAGAGACTACCCCATGCTGAGAACTGACAGGGCCAGATG 420  
QY 421 GTGCAACGAGCCATCAACCCCGCAACCTGAAACGCTGGGTGAAGGTGATCGAGAGAGAG 480  
Db 421 GTGCAACGAGCCATCAACCCCGCAACCTGAAACGCTGGGTGAAGGTGATCGAGAGAGAG 480

QY 481 GCCTTCAGCCCCGAGGTGATCCCAATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
DB 481 GCCTTCAGCCCCGAGGTGATCCCAATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCACAGCGCGCCATGAGATGCTGAAG 600  
DB 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCACAGCGCGCCATGAGATGCTGAAG 600  
QY 601 GACACCATCAACGAGGAGGCGCGCGAGTGGACCGCGTGACACCGCGTGACACGCGGCCCC 660  
DB 601 GACACCATCAACGAGGAGGCGCGCGAGTGGACCGCGTGACACCGCGTGACACGCGGCCCC 660  
QY 661 ATCGCCCCCGGCGAGATGCGCGAGCGCGCGGCGGACGACATCGCGGCGACACCAAGCAACC 720  
DB 661 ATCGCCCCCGGCGAGATGCGCGAGCGCGCGGCGGACGACATCGCGGCGACACCAAGCAACC 720  
QY 721 CTGCAAGAGCAGATCGCTGATGACCAAGCAACCCGCCATCCCGGTGGGCGACATCTAC 780  
DB 721 CTGCAAGAGCAGATCGCTGATGACCAAGCAACCCGCCATCCCGGTGGGCGACATCTAC 780  
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 840  
DB 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 840  
QY 841 CTGCAATCAAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
DB 841 CTGCAATCAAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
QY 901 ACCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
DB 901 ACCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
QY 961 GTGCAAGAACGCCAACCCCGACTGCAAGAACCATCTCGCGCGCTCTCGCGCGCGCGCGCG 1020  
DB 961 GTGCAAGAACGCCAACCCCGACTGCAAGAACCATCTCGCGCGCTCTCGCGCGCGCGCGCG 1020  
QY 1021 CTGGAAGAGATGATGACCGCGCTGCGCAAGGGCGGTGGGCGCGCGCGCGCGCGCGCG 1080  
DB 1021 CTGGAAGAGATGATGACCGCGCTGCGCAAGGGCGGTGGGCGCGCGCGCGCGCGCGCG 1080  
QY 1081 CTGGCCGAGGCGGATGAGCCAGGCGCAACCAAGCGTGATGATGACAGAGCAACTTCAAG 1140  
DB 1081 CTGGCCGAGGCGGATGAGCCAGGCGCAACCAAGCGTGATGATGACAGAGCAACTTCAAG 1140  
QY 1141 GGGCCCCCGGCGCATCTGTCAGTGTCTTCACTGGCGCAAGGAGGCGCACATCGCGCGCAAC 1200  
DB 1141 GGGCCCCCGGCGCATCTGTCAGTGTCTTCACTGGCGCAAGGAGGCGCACATCGCGCGCAAC 1200  
QY 1201 TGCCGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGCGGCGCAAGGAGGCGCACAGATGAAG 1260  
DB 1201 TGCCGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGCGGCGCAAGGAGGCGCACAGATGAAG 1260  
QY 1261 GACTGCACCGAGCGCGCGCAACTTCTCTGCGCAAGATCTGCGCGCACCAAGGCGCGC 1320  
DB 1261 GACTGCACCGAGCGCGCGCAACTTCTCTGCGCAAGATCTGCGCGCACCAAGGCGCGC 1320  
QY 1321 CCGCGCAACTTCTCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
DB 1321 CCGCGCAACTTCTCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
QY 1381 TTGAGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440  
DB 1381 TTGAGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440  
QY 1441 CTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAGTAA 1479  
DB 1441 CTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAGTAA 1479

RESULT 3  
AX468547  
LOCUS AX468547 1479 bp DNA linear PAT 16-Jul-2002

DEFINITION Sequence 67 from Patent WO0226209.  
ACCESSION AX468547  
VERSION AX468547.1 GI:21901377  
KEYWORDS  
SOURCE  
ORGANISM Human immunodeficiency virus 1 (HIV-1)  
Human immunodeficiency virus 1  
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate  
lentivirus group.  
REFERENCE  
AUTHORS O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singh,M., Ulmer,J. and Dubensky,T.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 67 04-APR-2002;  
CHIRON CORPORATION (US)  
FEATURES  
source  
1. 1479  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"

Query Match 99.6%; Score 1473.4; DB 6; Length 1479;  
Best Local Similarity 99.3%; Pred. No. 3.8e-150;  
Matches 1468; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
DB 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
DB 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
QY 121 CTGGAAGAGTTCGCGCTGGAACCCCGCGCTGCTGGAAGACCAAGCGGCTGCAAGCATC 180  
DB 121 CTGGAAGAGTTCGCGCTGGAACCCCGCGCTGCTGGAAGACCAAGCGGCTGCAAGCATC 180  
QY 181 ATCCGCCAGCTGCAACCCCGCGCTGCAAGACCGGCGAGGAGGCTGAAGAGCCTGTTCAAC 240  
DB 181 ATCCGCCAGCTGCAACCCCGCGCTGCAAGACCGGCGAGGAGGCTGAAGAGCCTGTTCAAC 240  
QY 241 ACCGTGGCCACCCCTGTACTGCGTGCACGAGAGATCGAGGTCCCGGCAACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCCCTGTACTGCGTGCACGAGAGATCGAGGTCCCGGCAACCAAGAGGCC 300  
QY 301 CTGGAACAAGATGAGGAGGAGGAGCAACAAGTGCCAGCAGAAAGATCCAGAGGCGCGCC 360  
DB 301 CTGGAACAAGATGAGGAGGAGGAGCAACAAGTGCCAGCAGAAAGATCCAGAGGCGCGCC 360  
QY 361 GCGGACAAAGGGGCAAGGTGAGCGCAACTACCCCATCTGTCAGAACTTGAGGGCGCAGATG 420  
DB 361 GCGGACAAAGGGGCAAGGTGAGCGCAACTACCCCATCTGTCAGAACTTGAGGGCGCAGATG 420  
QY 421 GTGCACCAAGGCATCAGCCCCCGCAACCTTGAACGCGCTGGTGAAGGTGATCGAGAGAA 480  
DB 421 GTGCACCAAGGCATCAGCCCCCGCAACCTTGAACGCGCTGGTGAAGGTGATCGAGAGAA 480  
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTTACCGCGCTGAGCGAGGGCGCGCACCCCGCAG 540  
DB 481 GCCTTCAGCCCCGAGGTGATCCCATGTTTACCGCGCTGAGCGAGGGCGCGCACCCCGCAG 540  
QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
DB 541 GACCTGAACACGATGTTGAACACCGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
QY 601 GACACCATCAACGAGGAGGCGCGCGAGTGGACCGCGTGACACCGCGTGACACGCGGCCCC 660  
DB 601 GACACCATCAACGAGGAGGCGCGCGAGTGGACCGCGTGACACCGCGTGACACGCGGCCCC 660  
QY 661 ATCGCCCCCGGCGAGATGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
DB 661 ATCGCCCCCGGCGAGATGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720

```
QY 721 CTGCAAGAGCAGATCGCCTGGATGACCAACCCCCCATCCCGTGGGCGACATCTAC 780
Db 721 CTGCAAGAGCAGATCGCCTGGATGACCAACCCCCCATCCCGTGGGCGACATCTAC 780
QY 781 AAGCGGTGATCATCTTGGGCTTGAACAAGATCGTGGGATGTACAGCCCCGTGACATC 840
Db 781 AAGCCSTGATCATCTTGGGCTTGAACAAGATCGTGGGATGTACAGCCCCGTGACATC 840
QY 841 CTGACATCAAGCAGAGGGCCCCCAAGAGCCCTTCCCGCACTACGTGAGACCGCTTCTCAAG 900
Db 841 CTGACATCAAGCAGAGGGCCCCCAAGAGCCCTTCCCGCACTACGTGAGACCGCTTCTCAAG 900
QY 901 ACCCTGCGCGCCGAGCAGAGCACCAGAGGTGAAGAACTGTGATGACCGACACCTGCTG 960
Db 901 ACCCTGCGCGCCGAGCAGAGCACCAGAGGTGAAGAACTGTGATGACCGACACCTGCTG 960
QY 961 GTGCAAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCCGGCGCCAGC 1020
Db 961 GTGCAAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCCGGCGCCAGC 1020
QY 1021 CTGGAAGAGATGATGACCGCCTGCGAAGGGCGGTGGCGGGCCCCAGCCACAAGGCCGCTG 1080
Db 1021 CTGGAAGAGATGATGACCGCCTGCGAAGGGCGGTGGCGGGCCCCAGCCACAAGGCCGCTG 1080
QY 1081 CTGGCCGAGGCGATGAGCCAGGCCAACACCAAGCGTGTATGATGAGAAAGAGCAACTTCAAG 1140
Db 1081 CTGGCCGAGGCGATGAGCCAGGCCAACACCAAGCGTGTATGATGAGAAAGAGCAACTTCAAG 1140
QY 1141 GGGCCCCGGCGCATCTGCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATGCGCCGCAAC 1200
Db 1141 GGGCCCCGGCGCATCTGCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATGCGCCGCAAC 1200
QY 1201 TGGCGGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACCAAGTGAAG 1260
Db 1201 TGGCGGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACCAAGTGAAG 1260
QY 1261 GACTGCACCCGAGCGCCAGGCCCAACTTCTTGGGCAAGATCTGCGCCAGCCACAAGGGCGCG 1320
Db 1261 GACTGCACCCGAGCGCCAGGCCCAACTTCTTGGGCAAGATCTGCGCCAGCCACAAGGGCGCG 1320
QY 1321 CCGCGCAACTTCTTGCAGAGCGCGCCGAGCCCAAGCGCCCCCGCGCGAGAGCTTCCGC 1380
Db 1321 CCGCGCAACTTCTTGCAGAGCGCGCCGAGCCCAAGCGCCCCCGCGCGAGAGCTTCCGC 1380
QY 1381 TTGAGAGAGACCAACCCCGCGCGCAAGAGAGAGCAAGAGCGCGAGACCTTGACCAAGC 1440
Db 1381 TTGAGAGAGACCAACCCCGCGCGCAAGAGAGAGCAAGAGCGCGAGACCTTGACCAAGC 1440
QY 1441 CTGAAGAGCCTGTTGGGCAACGACCCCTGAGCCAGTAA 1479
Db 1441 CTGAAGAGCCTGTTGGGCAACGACCCCTGAGCCAGTAA 1479
```

```
RESULT 4
AX455904
LOCUS AX455904 1479 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 20 from Patent WO0204493.
ACCESSION AX455904
VERSION AX455904.1 GI:21714896
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
AUTHORS Polynucleotides encoding antigenic hiv type c polypeptides,
TITLE Patent: WO 0204493-A 20 17-JAN-2002;
JOURNAL CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source
1. 1479
/organism="synthetic construct"
/mol_type="unassigned DNA"
```

/db\_xref="taxon:32630"  
/note="synthetic Gag coding sequence of HIV strain  
AF110965"

## ORIGIN

Query Match 98.9%; Score 1463; DB 6; Length 1479;  
Best local Similarity 99.3%; Pred. No. 5e-149;  
Matches 1469; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
QY 1 ATGGGCGCGCGCCAGCATCTGCGCGCGCGCAAGCTGAGCGCTTGAGCGCATCCGC 60
Db 1 ATGGGCGCGCGCCAGCATCTGCGCGCGCGCAAGCTGAGCGCTTGAGCGCATCCGC 60
QY 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGCGCGCGAG 120
Db 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGCGCGCGAG 120
QY 121 CTGGAAGAGTTCGCCCTGAACCCCGCGCTGCTGGAAGACCAAGAGGGCTGCAAGCATC 180
Db 121 CTGGAAGAGTTCGCCCTGAACCCCGCGCTGCTGGAAGACCAAGAGGGCTGCAAGCATC 180
QY 181 ATCCGCCAGCTGCACCCCGCGCTGCAGACCGGCGAGCGAGAGCTGAAGAGCCTGTTCAAC 240
Db 181 ATCCGCCAGCTGCACCCCGCGCTGCAGACCGGCGAGCGAGAGCTGAAGAGCCTGTTCAAC 240
QY 241 ACCGTGGCCACCCCTGTACTGCGTGACAGAGAGATCGAGGTCCGCGACCAAGAGGGCC 300
Db 241 ACCGTGGCCACCCCTGTACTGCGTGACAGAGAGATCGAGGTCCGCGACCAAGAGGGCC 300
QY 301 CTGGAACAAGATTCGAGAGAGAGCAACAAGTCCAGCAGAGATCCAGAGCGCGAGGCC 360
Db 301 CTGGAACAAGATTCGAGAGAGAGCAACAAGTCCAGCAGAGATCCAGAGCGCGAGGCC 360
QY 361 GCCGCAAGGCGCAAGGTGAGCGCAAGTACCCCATCGTGCAAGCTGACAGGGCCAGATG 420
Db 361 GCCGCAAGGCGCAAGGTGAGCGCAAGTACCCCATCGTGCAAGCTGACAGGGCCAGATG 420
QY 421 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAAGCGCTGGTGAAGTGAAGAGAGAAG 480
Db 421 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAAGCGCTGGTGAAGTGAAGAGAGAAG 480
QY 481 GCCTTACGCGCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCAG 540
Db 481 GCCTTACGCGCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCAG 540
QY 541 GACCTGAACAAGATGTTGAACACCGTGGCGGCCACCAAGCGCCCATGCAATGCTGAAG 600
Db 541 GACCTGAACAAGATGTTGAACACCGTGGCGGCCACCAAGCGCCCATGCAATGCTGAAG 600
QY 601 GACACCATCAACGAGAGGCGCGCGAGTGGGACCGGTCACCCCGGTGACCGCGGCCCC 660
Db 601 GACACCATCAACGAGAGGCGCGCGAGTGGGACCGGTCACCCCGGTGACCGCGGCCCC 660
QY 661 ATCGCCCCCGCCAGATGCGCGAGAGCCCCCGCGCGAGCGACATCGCCGCGACCAAGCACC 720
Db 661 ATCGCCCCCGCCAGATGCGCGAGAGCCCCCGCGCGAGCGACATCGCCGCGACCAAGCACC 720
QY 721 CTGCAAGAGAGATGCGCCTGATGACCAAGACCCCCCATCCCGGTGGGAGCATCTAC 780
Db 721 CTGCAAGAGAGATGCGCCTGATGACCAAGACCCCCCATCCCGGTGGGAGCATCTAC 780
QY 781 AAGCGGTGATCATCTTGGGCTTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 840
Db 781 AAGCGGTGATCATCTTGGGCTTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 840
QY 841 CTGACATCAAGCAGAGGGCCCCCAAGAGCCCTTCCCGCACTACGTGAGACCGCTTCTCAAG 900
Db 841 CTGACATCAAGCAGAGGGCCCCCAAGAGCCCTTCCCGCACTACGTGAGACCGCTTCTCAAG 900
QY 901 ACCCTGCGCGCGAGCAGAGCACCAGAGGTGAAGAACTGTGATGACCGACACCTGCTG 960
Db 901 ACCCTGCGCGCGAGCAGAGCACCAGAGGTGAAGAACTGTGATGACCGACACCTGCTG 960
```

QY	961	GTGCGAAGCGCCAAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCCGGCGCCAGC	1020
Db	961	GTGCGAAGCGCCAAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCCGGCGCCAGC	1020
QY	1021	CTGGAGGAGATGATGACCGCCTTGCCAGGGCGTGGGCGGGCCCCAGCCACCAAGGCTCGCGTG	1080
Db	1021	CTGGAGGAGATGATGACCGCCTTGCCAGGGCGTGGGCGGGCCCCAGCCACCAAGGCTCGCGTG	1080
QY	1081	CTGGCCGAGGCGGATGAGCCAGGCCAACCAACGCTGATGATGTCAGAAAGACAACTTCAAG	1140
Db	1081	CTGGCCGAGGCGGATGAGCCAGGCCAACCAACGCTGATGATGTCAGAAAGACAACTTCAAG	1140
QY	1141	GGCCCCCGGCGCATGCTCAAGTGCTTCAACTGCGGCAAGAGGGCCACATCGCCCCGAAC	1200
Db	1141	GGCCCCCGGCGCATGCTCAAGTGCTTCAACTGCGGCAAGAGGGCCACATCGCCCCGAAC	1200
QY	1201	TGCGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGCGGCAAGAGGGCCACCAATGAAG	1260
Db	1201	TGCGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGCGGCAAGAGGGCCACCAATGAAG	1260
QY	1261	GACTGCACCGAGCGCCAGGCCAACTTCTTGCGGCAAGATCTGGGCCAGCCACAAGGGCCG	1320
Db	1261	GACTGCACCGAGCGCCAGGCCAACTTCTTGCGGCAAGATCTGGGCCAGCCACAAGGGCCG	1320
QY	1321	CCGGGCAACTTCTTGCAAGAGCGGCCCCGAGCCACCGCCCCCGCGCGAGCTTCCGC	1380
Db	1321	CCGGGCAACTTCTTGCAAGAGCGGCCCCGAGCCACCGCCCCCGCGAGCTTCCGC	1380
QY	1381	TTTCGAGGAGACCAACCCCGGCGCAAGAGCAGAGAGCAAGAGCCGCGAGACCTGACCAAGC	1440
Db	1381	TTTCGAGGAGACCAACCCCGGCGCAAGAGCAGAGAGCAAGAGCCGCGAGACCTGACCAAGC	1440
QY	1441	CTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAGTAA 1479	
Db	1441	CTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAGTAA 1479	

[illegible]

|||||  
DB 781 TACAAGCGTGATCATCTGGGCTGAAACAAGTCGTGGCATGTACAGCCCGTGAGC 840  
QY 838 ATCTTGACATCAAGCAGGGGCCCCCAAGAGGCCCTTCCGCGACTACGTGAGCCGCTTCTTC 897  
DB 841 ATCTTGACATCAAGCAGGGGCCCCCAAGAGGCCCTTCCGCGACTACGTGAGCCGCTTCTTC 900  
QY 898 AAGACCTGCGCGCCGAGCAGAGCAACCCAGAGGTGAAGAACTGGATGACCGACACCCTG 957  
DB 901 AAGACCTGCGCGCCGAGCAGAGCAACCCAGAGGTGAAGAACTGGATGACCGACACCCTG 960  
QY 958 CTGGTGCAAGACGCCCAACCCCGACTGCAAGAACCTCTGCGCGCTCTGCGCCCGCGCGCC 1017  
DB 961 CTGGTGCAAGACGCCCAACCCCGACTGCAAGAACCTCTGAGGCGCTGGGCGCGCGCGCC 1020  
QY 1018 AGCCTGAGAGATGATGACCCGCTGCGAGGCGGTGGGCGCGCCCGCAAGAGCCCGC 1077  
DB 1021 ACCCTGAGAGATGATGACCCGCTGCGAGGCGGTGGGCGCGCCCGCAAGAGCCCGC 1080  
QY 1078 GTGCTGGCCGAGCGATGAGCCAGGCCCAACACC--AGCGTATGATGCAAGAGCAAC 1134  
DB 1081 GTGCTGGCCGAGCGCATGAGCCAGACCAACAGCGTGAACATCTGATGCAAGAGCAAC 1140  
QY 1135 TTCAAGGGGCCCCCGCGCATCGTCAAGTCTTCAACTGCGGCAAGAGGCGCACATCGCC 1194  
DB 1141 TTCAAGGGGCAACAGCGCATGTGAAGTCTTCAACTGCGGCAAGAGGCGCACATCGCC 1200  
QY 1195 CGCAACTGCGCGCGCCCCCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGCAACCAAG 1254  
DB 1201 CGCAACTGCGCGCGCCCCCGCAAGAGGCGTGTGAAGTGGCGCAAGAGGCGCAACCAAG 1260  
QY 1255 ATGAAGACTGCAACCGAGCGCCAGGCCCACTTCTGGGGCAAGATCTGGCCAGCCACAAG 1314  
DB 1261 ATGAAGACTGCAACCGAGCGCCAGGCCCACTTCTGGGGCAAGATCTGGCCAGCCACAAG 1320  
QY 1315 GGGCGCGCGCGCACTTCTGTGAGAGCGCGCCCGGACCCACCGCGCGCGCGAGAGC 1374  
DB 1321 GGGCGCGCGCGCACTTCTGTGAGAGCGCGCCCGGACCCACCGCGCGCGCGAGAGC 1380  
QY 1375 TTCCGCTTCGAGAGAACCAACCCCGCGCGCGCGAGAGCGAGAGCAAGCAACCTG 1434  
DB 1381 TTCCGCTTCGAGAGAACCAACCCCGCGCGCGCGAGAGCGAGAGCAAGCAACCTG 1440  
QY 1435 ACCAGCCTGAAGAGCGCTGTTCCGCAACGAGCCCTGAGCCAGTAA 1479  
DB 1441 ACCAGCCTGAAGAGCGCTGTTCCGCAACGAGCCCTGAGCCAGTAA 1485

RESULT 6  
AX455983 1491 bp DNA linear PAT 06-JUL-2002  
LOCUS AX455983  
DEFINITION Sequence 99 from Patent WO0204493.  
ACCESSION AX455983  
VERSION AX455983.1 GI:21714967  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 99 17-JAN-2002;  
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
FEATURES  
source location/Qualifiers  
1..1491

ORIGIN  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Gag\_TV2\_C\_Zaopt"

Query Match 87.1%; Score 1288.8; DB 6; Length 1491;

Best Local Similarity 93.0%; Pred. No. 3.1e-130;  
Matches 1389; Conservative 0; Mismatches 87; Indels 18; Gaps 3;  
QY 1 ATGGGCGCGCGCGCGAGCATCTCGCGCGCGGCAAGCTGGAACGCTGGAGCGCATCCGC 60  
DB 1 ATGGGCGCGCGCGCGCGAGCATCTCGCGCGCGGCAAGCTGGAACAGTGGAGAAATCCGC 60  
QY 61 CTGGCGCGCGCGCGCGCAAGAGTGTCTCATGATGAAGCACTGTGTGGCGAGCGCGAG 120  
DB 61 CTGGCGCGCGCGCGCGCGAGCAACTCATGCTGAAGCACTGTGTGGCGAGCGCGAG 120  
QY 121 CTGGAAGATTGCGCCCTGAACCCCGGCTGTGAGACCAAGGAGGCTGCAAGCAGATC 180  
DB 121 CTGAGCGGCTTCGCGGTGAACCCCGGCTGTGAGACCAAGGAGATCCGAGCTGTTCAAC 180  
QY 181 ATCCGCGAGTGCACCCCGGCTGCAGACCGGCGACCGAGAGATCCGAGCTGTTCAAC 240  
DB 181 ATCAAGCAGTGCAGCCCGGCTGCAGACCGGCGACCGAGAGATCCGAGCTGTTCAAC 240  
QY 241 ACCGTGGCCACCTGTACTGTGCGTGCAACGAGAAGATCGAGTCCGCGACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGTGCGTGCAACGAGAAGATCGAGTCCGCGACCAAGAGGCC 300  
QY 301 CTGGAACAAGTTCAGGAGGAGCAAGAACAGTGCAGCAAGAGATCCAGCGCGGAGGCC 360  
DB 301 CTGGAACAAGTTCAGGAGGAGCAAGAACAGTGCAGCAAGAGATCCAGCGCGGAGGCC 360  
QY 361 GCCGACAAGGCGAAGGTGAGGCCAAGACTACCCCATCGTGCAAGACCTGCAAGGCGCAGATG 420  
DB 361 GCCGACAAG--AAGGTAGGCCAAGACTACCCCATCGTGCAAGACCTGCAAGGCGCAGATG 417  
QY 421 GTGCAACCAAGGCCATCAGCCCCCGCACCCCTGAAAGCCTGGGTGAAGTGAAGAGAAG 480  
DB 418 GTGCAACCAAGGCCATCAGCCCCCGCACCCCTGAAAGCCTGGGTGAAGTGAAGAGAAG 477  
QY 481 GCCTTCAAGCCCGGAGGTGATCCCATGTGTCAACCGCCCTGAGCGAGGCGGCCACCCCGCAG 540  
DB 478 GCCTTCAAGCCCGGAGGTGATCCCATGTGTCAACCGCCCTGAGCGAGGCGGCCACCCCGCAG 537  
QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCCACCAAGCGCCCATGAGATGCTGAAG 600  
DB 538 GACCTGAACACGATGCTGAACACCGTGGGCGGCCACCAAGCGCCCATGAGATGCTGAAG 597  
QY 601 GACACCATCAACGAGGAGGCGCGCGAGTGGGACCGCGTGCACCCCGGTGCACGCGCGCGCCC 660  
DB 598 GACACCATCAACGAGGAGGCGCGCGAGTGGGACCGCGTGCACCCCGGTGCACGCGCGCGCCC 657  
QY 661 ATCGCCCCCGGCGCAGATGCGCGAGCCCGCGCGCAGGACATCGCGCGGCAACCAAGCAGACC 720  
DB 658 GTGGCCCCCGGCGCAGATGCGCGAGCCCGCGCGCAGGACATCGCGCGGCAACCAAGCAGACC 717  
QY 721 CTGCAAGAGCAGATGCGCTGTGATGACCAAGCAACCCCGCATCCCGGTGGGCGACATCTAC 780  
DB 718 CTGCAAGAGCAGATGCGCTGTGATGACCAAGCAACCCCGCATCCCGGTGGGCGACATCTAC 777  
QY 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCAGATGTACAGCCCGGTGAGCATC 840  
DB 778 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCAGATGTACAGCCCGGTGAGCATC 837  
QY 841 CTGGAATCAAGCAGGCGCGCGCGCGCGCATCTCGCGCATCTGAGCCGCTTCTTCAAG 900  
DB 838 CTGGAATCAAGCAGGCGCGCGCGCGCGCGCATCTCGCGCATCTGAGCCGCTTCTTCAAG 897  
QY 901 ACCCTGCGCGCGCGAGCAGAGCAACCCAGAGGTGAAGAACTGAGATGACCGACACCTGCTG 960  
DB 898 ACCCTGCGCGCGCGAGCAGAGCAACCCAGAGGTGAAGAACTGAGATGACCGACACCTGCTG 957  
QY 961 GTGCAAGACGCCCAACCCCGACTGCAAGAACCATCTCTGCGCGCTCTCGGCGCGCGCGCAGC 1020  
DB 958 GTGCAAGACGCCCAACCCCGACTGCAAGAACCATCTCTGCGCGCGCTGCGGCGCGCGCAGC 1017  
QY 1021 CTGGAAGAGATGATGACCGCGCTGCGAGGCGCGTGGGCGCGCGCGCGCAGCAGAGGCCGCGTG 1080

Db	1018	CTGGAGGAGATGATGACCGGCTGTGGCAGGGCGTGGGGCGGGCCCGCAGCCACCAAGGGCCCGCGTG	1077
Qy	1081	CTGGCCGAGGGCGATGAGCCAGGC---CAACACACGCGTGATGATGCAGAAAGCAACTTC	1137
Db	1078	CTGGCCGAGGGCCATGAGCCAGGCCAACACACACGCGTGATGATCCAGAAAGCAACTTC	1137
Qy	1138	AAGGGCCCCCGGCGCATGCTCAAGTGCTTCAACTGCGGCAAGAGAGGGCCACATCGCCCGC	1197
Db	1138	AAGGGCCCCCGCGCGCGCGCGTGAAGTGCTTCAACTGCGGCGCGGAGGGCCACATCGCCCGC	1197
Qy	1198	AACTGCGCGCGCGCGCGCGCGCAAGAAAGGGCTGCTGGAAGTGCGGCAAGAGGGCCACCAAGATG	1257
Db	1198	AACTGCGCGCGCGCGCGCGCGCAAGAGCGGGCTGCTGGAAGTGCGGCAAGAGGGCCACCAAGATG	1257
Qy	1258	AAGGACTGCACCGAGCGCCAGGCGCAACTTCTTGGGGCAAGATTTGGCCACGCCACAAAGGGC	1317
Db	1258	AAGGACTGCACCGAGCGCCAGGCGCAACTTCTTGGGGCAAGATTTGGCCACGCCACAAAGGGC	1317
Qy	1318	CGGCCCCGGGCAACTTCTCTGCAGAGGCGGCGCCGAGGCCACCGGCCCCCGCGCGAGAGCTTC	1377
Db	1318	CGGCCCCGGGCAACTTCTCTGCAGAGGCGGCGCCGAGGCCACCGGCCCCCGCGCGAGAGCCACC	1377
Qy	1378	CGGCTTCGAGG-----AGACCAACCCCGGCGCAAGAGCAGAGAGCAGAGACCGC	1425
Db	1378	GGCCCCCCCCCGCGAGAGAGCTTCAAGTTCAAGAGAGACCCCCCAAGCAGAGGCCCAAGAACCGC	1437
Qy	1426	GAGACCCCTGACCAAGCCCTGAAGAGCCTGTTGGGCAACGAGACCCCTGAGCCAGTAA	1479
Db	1438	GAGCCCTTGACCAAGCCCTGAAGAGCCTGTTGGGCAACGAGACCCCTGAGCCAGTAA	1491

RESULT 7	LOCUS	AX455935	AX455935	1494 bp	DNA	linear	PAT 06-JUL-2002
DEFINITION	Sequence 51 from Patent WO0204493.						
ACCESSION	AX455935						
VERSION	AX455935.1						
KEYWORDS	GI:21714919						
SOURCE	synthetic construct						
ORGANISM	synthetic construct						
REFERENCE	other sequences; artificial sequences.						
AUTHORS	zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.						
TITLE	Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof						
JOURNAL	Patent: WO 0204493-A 51 17-JAN-2002;						
FEATURES	CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)						
source	Location/Qualifiers						
	1. .1494						
	/organism="synthetic construct"						
	/mol_type="unassigned DNA"						
	/db_xref="taxon:32630"						
	/note="HIV Type C Gag optimized"						
ORIGIN							
Query Match	87.1%;	Score 1288;	DB 6;	Length 1494;			
Best Local Similarity	92.9%;	Pred. No. 3.7e-130;					
Matches 1393;	Conservative 0;	Mismatches 80;	Indels 27;	Gaps 3;			
QY	1 ATGGGCGCCCGCGCCAGCATCTGCGCGCGGCAAGCTGACGCGCTGGAGCGCATCCGC	60					
DB	1 ATGGGCGCCCGCGCCAGCATCTGAGCGGCGCAAGCTGACCAAGTGGAGCGCATCCGC	60					
QY	61 CTGCGCCCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGCGCAGCGCGAG	120					
DB	61 CTGCGCCCGCGCGCAAGAGCACTACATGTGAAGCACTGTGTGGCGCAGCGCGAG	120					
QY	121 CTGAGAGAAGTTGCGCCTGAACCCCGGCGCTGTGAGACCAAGGAGGCTGCAAGCAGATC	180					
DB	121 CTGAGAGCGCTTGCCTGTGAACCCCGGCGCTGTGAGAGACCAAGGAGGCTGCAAGCAGATC	180					
QY	181 ATCCGCCAGCTGCACCCCGCCTGACAGACCGGCGAGGAGAGCTGAAGCCCTGTTCAAC	240					

Db	181	ATCAAGCAGCTGCAGCCCGCCCTGCAGACC	GGGCAACCGAGGAGCTGGCAGCCTGTTCAAC	240
QY	241	ACCGTGGCCACCCCTGTACTTGGGTGCACG	AAGATCGAGTCCGCGACACCAAGGAGGCC	300
Db	241	ACCGTGGCCACCCCTGTACTTGGGTGCACG	AAGGCGCATCGAGGTGCGGACACCAAGGAGGCC	300
QY	301	CTGGACAAGATCGAGGAGGAGCAAA	CAAGTGGCCAGCAGAAGATCCAGCAGGCCGAGGCC	360
Db	301	CTGGACAAGATCGAGGAGGAGCAAA	CAAGTGGCCAGCAGAAGGCCCCAGCAGGCCCAAGGCC	360
QY	361	GCCGACAAGGGCGAAGGTGAGCCAGAA	CTACCCCATCTGTGAGAACTTGAGGGCCAGATG	420
Db	361	GCCGACGAG---AAGGTAGCCAGAACT	ATCCCATCTGTGAGAAACGCCAGGGCCAGATG	417
QY	421	GTGCACCAAGGCCATCAGCCCCCGCAC	ACCCTGAACGCCCTGGGTGAAGGTGATCGAGGAGAA	480
Db	418	GTGCACCAAGGCCATCAGCCCCCGCAC	ACCCTGAACGCCCTGATCAAGGTGATCGAGGAGAA	477
QY	481	GCGTTCAAGCCCGGAGGTGATCCCAT	GTGTCACCGCCCTGAGCGAGGGCGGCCACCCCCAG	540
Db	478	GCGTTCAAGCCCGGAGGAGATCCCAT	GTGTCACCGCCCTGAGCGAGGGCGGCCACCCCCAG	537
QY	541	GACCTGAACACGATGTTGAACAC	CGTGGGCGGCCACCAAGGCGCCATGCAATGCTGAAG	600
Db	538	GACCTGAACACGATGCTGAACAC	CGTGGGCGGCCACCAAGGCGCCATGCAATGCTGAAG	597
QY	601	GACACCATCAACGAGGAGCGCGCGAG	TGGGACCGCGTGCACCCCGTGACAGCGCGGCCCC	660
Db	598	GACACCATCAACGAGGAGCGCGCGAG	TGGGACCGCGTGCACCCCGTGACAGCGCGGCCCC	657
QY	661	ATCGCCCCCGCCAGATGCGCGAG	CCCCCGCGGACGCAATCGCCCGCACCAACAGCACC	720
Db	658	GTGGCCCCCGCCAGATGCGCGAG	CCCCCGCGGACGCAATCGCCCGCACCAACAGCACC	717
QY	721	CTGCAGGAGCAGATCGCCTTGATGAC	CAGCAACCCCCCATCCCGTGGCGCATCTTAC	780
Db	718	CTGCAGGAGCAGATCGCCTTGATGAC	CAGCAACCCCCCATCCCGTGAAGACATCTTAC	777
QY	781	AAGCGGTGATCATCCTGGGCTGAACA	AGATCGTGGCATGTACAGCCCCGTGAGCATC	840
Db	778	AAGCGGTGATCATCCTGGGCTGAACA	AGATCGTGGCATGTACAGCCCCGTGAGCATC	837
QY	841	CTGGACATCAAGCAGGGGCCCAAGG	AGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG	900
Db	838	CTGGACATCAAGCAGGGGCCCAAGG	AGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG	897
QY	901	ACCCTGCGCGCGAGCAGAGCA	CCCAAGAGGTGAAGAACTGGATGACCGACACCCCTGCTG	960
Db	898	ACCCTGCGCGCGAGCAGAGCA	CCCAAGAGGTGAAGAACTGGATGACCGACACCCCTGCTG	957
QY	961	GTGCAGAAACGCCAAACCCCGA	CTGCAGACCATCTGCGCGCTCTCGGCCCGGGCGGCAGC	1020
Db	958	GTGCAGAAACGCCAAACCCCGA	CTGCAGACCATCTGCGCGCTCTGGGCCCGGGCGGCAGC	1017
QY	1021	CTGGAGGAGATGATGACCGCCTG	CCAGGGCGTGGGCGGCCAGCCACAAAGGCCCGCGTG	1080
Db	1018	CTGGAGGAGATGATGACCGCCTG	CCAGGGCGTGGGCGGCCAGCCACAAAGGCCCGCGTG	1077
QY	1081	CTGGCCGAGGCGATGAGCCAGGCC	CAACACCAAGCGTGATGATGCAGAAAGCAACTTCAAG	1140
Db	1078	CTGGCCGAGGCGATGAGCCAGGCC	CAACACCAAGCAATCTGTGTGACGCGCAACAATTCAAG	1137
QY	1141	GCGCCCCCGCGCATCTGTCAGT	GTCTTCAACTGCGGCAAGGTGGCCACATCGCCCGCAAC	1200
Db	1138	GCGCAACACCGCATCATCAAGT	GTCTTCAACTGCGGCAAGGTGGCCACATCGCCCGCAAC	1197
QY	1201	TGCGCGCGCCCCCGCAAGAGGG	CTGTGAAGTGCAGCAAGAGGGCCACAGATGAAG	1260
Db	1198	TGCGCGCGCCCCCGCAAGAGGG	CTGTGAAGTGCAGCAAGAGGGCCACAGATGAAG	1257
QY	1261	GACTGCACCGAGCGCCAGGCCAA	CTTCTGTGGCAAGATCTGGCCAGGCCCAAGGGCGCG	1320
Db	1258	GACTGCACCGAGCGCCAGGCCAA	CTTCTGTGGCAAGATCTGGCCAGGCCCAAGGGCGCG	1317

QY 1321 CCGGCACTTCTGTGAGAGCGGCC-----CCGAGCCACCGCC 1359  
|||||  
Db 1318 CCGGCACTTCTGTGAGAACCGGCCCCGAGCCACCGCCCCCGGAGCCACCGCC 1377  
QY 1360 CCCCCCGCGAGAGCTTCCGCTTGAGAGAGACACCCCGGCGCAAGAGAGAGCAAG 1419  
|||||  
Db 1378 CCCCCCGCGAGAGCTTCCGCTTGAGAGAGACACCCCGGCGCAAGAG---AAG 1434  
QY 1420 GACCGCGAGAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
|||||  
Db 1435 GAGCGCGAGAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494

RESULT 8  
AX455888 1509 bp DNA linear PAT 06-JUL-2002  
LOCUS AX455888  
DEFINITION Sequence 4 from Patent WO0204493.  
ACCESSION AX455888  
VERSION AX455888.1 GI:21714881  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1 other sequences; artificial sequences.  
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
JOURNAL Patent: WO 0204493-A 4 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES  
source 1.1509  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="synthetic Gag of HIV strain AF110967"

ORIGIN  
Query Match 86.3%; Score 1276.8; DB 6; Length 1509;  
Best Local Similarity 92.5%; Pred. No. 6e-129;  
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;

QY 1 ATGGGCGCCCGCGCAGCATCTGCGCGCGCGCAAGCTTGAGCGCTGGAGCGCATCCGC 60  
|||||  
Db 1 ATGGGCGCCCGCGCAGCATCTGCGCGCGCGCAAGCTTGAGCGCTGGAGAGATCCGC 60  
QY 61 CTGCGCCCGCGCGCAAGAGTGCTACATGATGAGCACCTGTGTGGCGCAGCGCGAG 120  
|||||  
Db 61 CTGCGCCCGCGCGCAAGAGACTACATGCTGAGGACCTGTGTGGCGCAGCGCGAG 120  
QY 121 CTGAGAGAAGTTGCGCTGAACCCCGGCTGTGAGAGACGAGGGCTGCAAGCATC 180  
|||||  
Db 121 CTGAGAGGCTTGCCTGAACCCCGGCTGTGAGAGACGCGAGGGCTGCAAGCATC 180  
QY 181 ATCCGCGAGCTGACACCCCGCTGAGACCGGCGAGAGAGCTGAAGCCTGTCAAC 240  
|||||  
Db 181 ATGAAGCAGCTGACACCCCGCTGAGACCGGCGAGAGAGCTGCGCAGCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGACAGAGAGATCGAGTCCGCGACACCAAGAGGCC 300  
|||||  
Db 241 ACCGTGGCCACCTGTACTGCGTGACAGCGCGGCGATCGAGTCCGCGACACCAAGAGGCC 300  
QY 301 CTGAGACAAGATCGAGAGAGAGAGACAAGTGCAGCAGAGAGATCCAGCGCGAGGCC 360  
|||||  
Db 301 CTGAGACAAGATCGAGAGAGAGAGACAAGTCCAGCAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCCGACAAGGGCAAGGTGAGCGCAACTACCCCATCTGTGACAACCTGCAAGGGCCAGATG 420  
|||||  
Db 361 GCCGAC--GGCAAGGTGAGCGCAACTACCCCATCTGTGACAACCTGCAAGGGCCAGATG 417  
QY 421 GTGCAACGAGGCGCATGAGCCCCCGCACCTGAAAGCCTGGGTGAAGGTGATCGAGAGAG 480  
|||||  
Db 418 GTGCAACGAGGCGCATGAGCCCCCGCACCTGAAAGCCTGGGTGAAGGTGATCGAGAGAG 477

QY 481 GCCTTACGCCCCGAGGTGATCCCATGTTCAACGCCCCGTGAGCGAGGGCGCCACCCCCAG 540  
|||||  
Db 478 GCCTTACGCCCCGAGGTGATCCCATGTTCAACGCCCCGTGAGCGAGGGCGCCACCCCCAG 537  
QY 541 GACCTGAACAGATGTTGAACACCGTGGGGCGCCACAGGCGCCATGACATGCTGAAG 600  
|||||  
Db 538 GACCTGAACAGATGTTGAACACCGTGGGGCGCCACAGGCGCCATGACATGCTGAAG 597  
QY 601 GACACCATCAAGAGAGAGCGCGGAGTGGGACCGGTGCAACCCCGTGACAGCGCGGCC 660  
|||||  
Db 598 GACACCATCAAGAGAGAGCGCGGAGTGGGACCGGTGCAACCCCGTGACAGCGCGGCC 657  
QY 661 ATGCCCCCGGCGAGATGCGCGAGAGCGCGGCGAGCGACATCGCGGCAACGACACC 720  
|||||  
Db 658 GTGGCCCCCGGCGAGATGCGCGAGAGCGCGGCGAGCGACATCGCGGCGCCACGACACC 717  
QY 721 CTGCAAGAGAGATCGCTGTGATGACCAAGACACCCCGCATCCCGGTGGCGCATCTAC 780  
|||||  
Db 718 CTGCAAGAGAGATCGCTGTGATGACCAAGACACCCCGCATCCCGGTGGCGCATCTAC 777  
QY 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGGTGAGCATC 840  
|||||  
Db 778 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGGTGAGCATC 837  
QY 841 CTGCAATCAAGCAGAGAGCGCCCAAGAGAGCGCTCCGCGACTACGTGAGACCGCTTCAAG 900  
|||||  
Db 838 CTGCAATCGCCAGAGAGCGCCCAAGAGAGCGCTCCGCGACTACGTGAGACCGCTTCAAG 897  
QY 901 ACCCTGCGCGCGAGCAGAGAGACCCAGAGAGTGAAGAACTGATGACCGACCCCTGCTG 960  
|||||  
Db 898 ACCCTGCGCGCGAGCAGAGAGACCCAGAGAGTGAAGAACTGATGACCGAGACCCCTGCTG 957  
QY 961 GTGCAAGACCGCCAAACCCCGCACTGCAAGACCATCTGCGCGCTCTCGGCGCGCGCAGC 1020  
|||||  
Db 958 GTGCAAGACCGCCAAACCCCGCACTGCAAGACCATCTGCGCGCTCTCGGCGCGCGCAGC 1017  
QY 1021 CTGAGAGAGATGATGACCGCTGCGAGAGGCGTGGGCGCGCCAGCCACAGAGCGCGTG 1080  
|||||  
Db 1018 CTGAGAGAGATGATGACCGCTGCGAGAGGCGTGGGCGCGCCAGAGCGCGCGTG 1077  
QY 1081 CTGCGCGAGGCGATGAGCGGCGCCAAAC---AGCGTATGATGCAAGAGCAACTTC 1137  
|||||  
Db 1078 CTGCGCGAGGCGATGAGCGGCGCCAAACAGCGTGAACATCATGATGCAAGAGCAACTTC 1137  
QY 1138 AAGGGCCCCCGCGCATCGTCAAGTCTTCAACTGCGGCAAGAGAGGGCCACATCGCCAG 1197  
|||||  
Db 1138 AAGGGCCCCCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGAGGGCCACATCGCCAG 1197  
QY 1198 AACTGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCAAGATG 1257  
|||||  
Db 1198 AACTGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCAAGATG 1257  
QY 1258 AAGGACTGACCGAGCGCGCGCAACTTCTGGGCAAGATCTGGCCCAAGAGGGC 1317  
|||||  
Db 1258 AAGGACTGACCGAGCGCGCGCAACTTCTGGGCAAGATCTGGCCCAAGAGGGC 1317  
QY 1318 CGCCCCGCAACTTCTGCAAGAGCG-----CCCCAGGCCACCGCC 1359  
|||||  
Db 1318 CGCCCCGCAACTTCTGCAAGAGCGAGCGAGCGCGCGCGCCACCGTGTCCACCGCC 1377  
QY 1360 CCCCCCGCGAGAGCTTCCGCTTGAGAGAGACACCCCGGCGCAAGAGAGAGAGAG 1419  
|||||  
Db 1378 CCCCCCGCGAGAGCTTCCGCTTGAGAGAGACACCCCGCGCGCAAGAGAGAGAGAG 1437  
QY 1420 G-----ACCGGAGAGCCCTGACCAAGCCTGGAAGAGCCTGTTGGCAACGACCCC 1467  
|||||  
Db 1438 GACCGGAGCGCTTACCGGAGCGCGCTGACCGCGCTGCGCAGCGCTGTTGGCAGCGGCC 1497  
QY 1468 CTGAGCCAGTAA 1479  
|||||  
Db 1498 CTGAGCCAGTAA 1509

RESULT 9  
LOCUS AX468544 1509 bp DNA linear PAT 16-JUL-2002  
DEFINITION Sequence 64 from Patent WO0226209.  
ACCESSION AX468544  
VERSION AX468544.1 GI:21901374  
KEYWORDS  
SOURCE Human immunodeficiency virus 1 (HIV-1)  
ORGANISM Human immunodeficiency virus 1  
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate  
Lentivirus group.  
REFERENCE 1  
AUTHORS O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singh,M., Ulmer,J. and Dubensky,T.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 64 04-APR-2002;  
CHIRON CORPORATION (US)  
FEATURES  
SOURCE location/Qualifiers  
1..1509  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"  
ORIGIN

Query Match 86.2%; Score 1275.2; DB 6; Length 1509;  
Best Local Similarity 92.5%; Pred. No. 8.9e-129;  
Matches 1398; Conservative 0; Mismatches 78; Indels 36; Gaps 4;

QY 1 ATGGGCGCCCGCCGACATCTCCGCGCGCGGCAAGCTGGACGCGCTGGAGCGCATCCG 60  
DB 1 ATGGGCGCCCGCCGACATCTCCGCGCGCGGCAAGCTGGAGCTGGACAACTGGAGAAATCCG 60  
QY 61 CTGGCGCCCGCGCGCAAGAGTCTACATGATGAAGCACTGTGTGGGCCAGCGCCGAG 120  
DB 61 CTGGCGCCCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGGCCAGCGCCGAG 120  
QY 121 CTGGAAGAAGTTCGCTTGAACCCCGGCTGTGAGACCAAGGAGCTGCAAGCATC 180  
DB 121 CTGGAAGGCTTCGCTTGAACCCCGGCTGTGAGACCGGAGGAGCTGCAAGCATC 180  
QY 181 ATCCGCCAGCTGCAACCCCGCTGCAAGACCGGCAAGGAGAGCTGAAGACCTGTTCAAC 240  
DB 181 ATGAAGCAGCTGCAACCCCGCTGCAAGACCGGCAAGGAGAGCTGCGACGCTGTCAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAAGAGATGAGGTCCGCGCAACCAAGAGGCGC 300  
DB 241 ACCGTGGCCACCTGTACTGCGTGCAAGGATGAGGTCCGCGCAACCAAGAGGCGC 300  
QY 301 CTGGAACAAGATCGAAGAGAGAGCAAGTCCCAAGAGAGATCCAGAGGCGCAAGGAG 360  
DB 301 CTGGAACAAGATCGAAGAGAGAGCAAGTCCCAAGAGAGATCCAGAGGCGCAAGGAG 360  
QY 361 GCGCAACAAGGCAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGTGAGGCGCAGATG 420  
DB 361 GCGCAAC--GCGAAGGTGAGCCAGAACTACCCCATCTGTGAGAACTGTGAGGCGCAGATG 417  
QY 421 GTGCAACGAGGCAATCAACCCCGCAACCTGAAGCCTGGGTGAAGGTGATGAGAGAAAG 480  
DB 421 GTGCAACGAGGCAATCAACCCCGCAACCTGAAGCCTGGGTGAAGGTGATGAGAGAAAG 480  
QY 418 GTGCAACGAGGCAATCAACCCCGCAACCTGAAGCCTGGGTGAAGGTGATGAGAGAAAG 477  
DB 418 GTGCAACGAGGCAATCAACCCCGCAACCTGAAGCCTGGGTGAAGGTGATGAGAGAAAG 477  
QY 481 GCGTTACGCGCCGAGGTGATCCCAATGTTCAACGCGCTGAGGAGGCGCCACCCCGCAG 540  
DB 478 GCGTTACGCGCCGAGGTGATCCCAATGTTCAACGCGCTGAGGAGGCGCCACCCCGCAG 537  
QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCAAGGCGCGCATGAGATGCTGAAG 600  
DB 538 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCAAGGCGCGCATGAGATGCTGAAG 597  
QY 601 GACACCATCAACGAGAGGCGCGCGAGGTGGACCGGCTGACCCCGTGCAGCGCGCGCC 660  
DB 598 GACACCATCAACGAGAGGCGCGCGAGGTGGACCGGCTGACCCCGTGCAGCGCGCGCC 657

QY 661 ATGCGCCCGCGCGCAGATGCGCGAGCGCCCGCGCGCAGCGACATCGCGCGCACCAAGCAGCACC 720  
DB 658 GTGGCCCCCGGCGCAGATGCGCGACCGCCCGCGCGCAGCGACATCGCGCGCGCACCAAGCAGCACC 717  
QY 721 CTGCAAGAGCAGATGCGCTGTGATGACCAAGCAACCCCGCATCCCGGTGGGCGACATCTAC 780  
DB 718 CTGCAAGAGCAGATGCGCTGTGATGACCAAGCAACCCCGCATCCCGGTGGGCGACATCTAC 777  
QY 781 AAGCGGTGATCATCTGTGGGCTGAAACAAGATCTGCGGATGTACAGCCCGTGAGCATC 840  
DB 778 AAGCGGTGATCATCTGTGGGCTGAAACAAGATCTGCGGATGTACAGCCCGTGAGCATC 837  
QY 841 CTGCAATCAAGCAGGCGCCCAAGAGCGCTTCCGCGACTACGTGAGACCGCTTCTTCAAG 900  
DB 838 CTGCAATCCGCGCAGGCGCCCAAGAGCGCTTCCGCGACTACGTGAGACCGCTTCTTCAAG 897  
QY 901 ACCCTGCGCGCGCAGCAGAGCAACCGAGGTGAAGAACTGATGACCGACACCTGCTG 960  
DB 898 ACCCTGCGCGCGCAGCAGGCGCACCGACGTAAGAACTGATGACCGAGACCTGCTG 957  
QY 961 GTGCAAGACGCAACCCCGACTGCAAGACCATCTGCGGCTTCGCGCCCGCGCGCAGC 1020  
DB 958 GTGCAAGACGCAACCCCGACTGCAAGACCATCTGCGGCTTCGCGCCCGCGCGCAGC 1017  
QY 1021 CTGGAAGAGATGATGACCGCTGCGCAGGCGCGTGGCGCGCCCGCAAGAGCGCGCTG 1080  
DB 1018 CTGGAAGAGATGATGACCGCTGCGCAGGCGCGTGGCGCGCCCGCAAGAGCGCGCTG 1077  
QY 1081 CTGGCGCAGGCGATGAGCGCAGCGCAACCC--AGCGTATGATGAGAGAGCAACTTC 1137  
DB 1078 CTGGCGCAGGCGATGAGCGCAGCGCAACCGTGAACATCATGATGAGAGAGCAACTTC 1137  
QY 1138 AAGGCGCCCGCGCGCATCTGTAAGTCTTCAACTGCGGCAAGAGGCGCACATCGCGCC 1197  
DB 1138 AAGGCGCCCGCGCGCAACGTCAAGTCTTCAACTGCGGCAAGAGGCGCACATCGCGCAAG 1197  
QY 1198 AACTGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGCAGAGAGGCGCACAGATG 1257  
DB 1198 AACTGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGCAGAGAGGCGCACAGATG 1257  
QY 1258 AAGACTGCAACGAGCGCGCAACTTCTGCGCAAGATCTGCGCGCAAGAGGCGCACAGGCG 1317  
DB 1258 AAGACTGCAACGAGCGCGCAACTTCTGCGCAAGATCTGCGCGCAAGAGGCGCACAGGCG 1317  
QY 1318 CGCGCGCGCAACTTCTGAGAGCG-----CGCGAGCGCGCGCGC 1359  
DB 1318 CGCGCGCGCAACTTCTGAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1377  
QY 1360 CCGCGCGCGGAGGCTTCCGCTTGAAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1419  
DB 1378 CCGCGCGCGGAGGCTTCCGCTTGAAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1437  
QY 1420 G-----ACCGGAGACCTTGACCAAGCTTGAAGAGCTTTCGCAACGACCCC 1467  
DB 1438 GACCGGAGCGCGCTACCGCGAGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1497  
QY 1468 CTGAGCCAGTAA 1479  
DB 1498 CTGAGCCAGTAA 1509

RESULT 10  
LOCUS AX468548 1509 bp DNA linear PAT 16-JUL-2002  
DEFINITION Sequence 68 from Patent WO0226209.  
ACCESSION AX468548  
VERSION AX468548.1 GI:21901378  
KEYWORDS  
SOURCE Human immunodeficiency virus 1 (HIV-1)  
ORGANISM Human immunodeficiency virus 1  
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate  
Lentivirus group.  
REFERENCE 1

AUTHORS O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singh,M., Ulmer,J. and Dubensky,T.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 68 04-APR-2002;  
CHIRON CORPORATION (US)

FEATURES  
source 1. .1509  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"

ORIGIN

Query Match 86.1%; Score 1273.6; DB 6; Length 1509;  
Best Local Similarity 91.8%; Pred. No. 1.3e-128;  
Matches 1388; Conservative 12; Mismatches 76; Indels 36; Gaps 4;

QY 1 ATGGCGCCCGCCAGCATCTGCGCGCGCAAGCTGACGCGCTGGAGCGCATCCG 60  
DB 1 ATGGCGCCCGCCAGCATCTGCGCGCGCAAGCTGACGCGCTGGAGCGCATCCG 60  
QY 61 CTGCGCCCGCGCGCAAGAGTGTACATGATGAAGCACCTGTGTGGCCAGCCGAG 120  
DB 61 CTGCGCCCGCGCGCAAGAGCACTACATGCTGAAGCACCTGTGTGGCCAGCCGAG 120  
QY 121 CTGGAAGAAGTGGCCCTGAACCCCGGCTGTGTGAGAGCCAGCGAGGCTGCAAGCATC 180  
DB 121 CTGGAAGGCTTGGCCCTGAACCCCGGCTGTGTGAGAGCCAGCGAGGCTGCAAGCATC 180  
QY 181 ATCCGCGAGCTGCAACCCGCTGTGACAGCCGCGAGAGAGCTGAAGAGCTGTCAAC 240  
DB 181 ATGAAGCAGCTGCAACCCGCTGTGACAGCCGCGAGAGAGCTGCGAGCTGTACAAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGTGACAGAAAGATCGAGTCCGCGACACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGCGTGTGACAGCCGCGAGATCGAGTCCGCGACACCAAGAGGCC 300  
QY 301 CTGGAACAAGATCGAGAGAGCAAAAGTGTGCGAGCAAGAGATCCAGAGCGCGAGGCC 360  
DB 301 CTGGAACAAGATCGAGAGAGCAAAAGTGTGCGAGCAAGAGATCCAGAGCGCGAGGCC 360  
QY 361 GCCGACAAGGGCAAGGTGAGCGCAACTACCCCATCTGTGAGAACTGTGAGGGCCAGATG 420  
DB 361 GCCGAC---GGCAAGGTGAGCGCAACTACCCCATCTGTGAGAACTGTGAGGGCCAGATG 417  
QY 421 GTGCAACGAGGCAATCAGCCCGCGACCTGAAGCGCTGGTGAAGGTGATCGAGGAAG 480  
DB 418 GTGCAACGAGGCAATCAGCCCGCGACCTGAAGCGCTGGTGAAGGTGATCGAGGAAG 477  
QY 481 GCCTTCAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCACCCCCAG 540  
DB 478 GCCTTCAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCACCCCCAG 537  
QY 541 GACCTGAACAGATGTTGAACAACCGTGGCGGCGCACAGCGCCATGACATGCTGAAG 600  
DB 538 GACCTGAACAGATGTTGAACAACCGTGGCGGCGCACAGCGCCATGACATGCTGAAG 597  
QY 601 GACACCATCAAGAGAGGCGCGAGTGTGAGCGCGGTGACCCCGTGCAGCGCGGCCCC 660  
DB 598 GACACCATCAAGAGAGGCGCGAGTGTGAGCGCGGTGACCCCGTGCAGCGCGGCCCC 657  
QY 661 ATCGCCCCCGCGAGATGCGCGAGCCCGCGCGAGCATCGCGCGCGCACAGCACCC 720  
DB 658 GTGGCCCCCGCGAGATGCGCGAGCCCGCGCGAGCATCGCGCGCGCACAGCACCC 717  
QY 721 CTGCAAGAGAGATGCGCTGATGACCAACCCCGCATCCCGTGGGCGAGATCTAC 780  
DB 718 CTGCAAGAGAGATGCGCTGATGACCAACCCCGCATCCCGTGGGCGAGATCTAC 777  
QY 781 AAGCGGTGATCATCTGGGCGTGAACAAGATGTGCGGATGTACAGCCCGTGAAGCATC 840  
DB 778 AAGCGGTGATCATCTGGGCGTGAACAAGATGTGCGGATGTACAGCCCGTGAAGCATC 837  
QY 841 CTGACATCAAGAGAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAGCCGCTTCAAG 900

DB 838 CTGACATCCGCGAGGCGCCCAAGAGAGCCCTCCGCACTACGTGAGCCGCTTCAAG 897  
QY 901 ACCCTGCGCGCGAGCAAGACACCCAGAGGTGAAGAATGTGATGACCGACACCTGCTG 960  
DB 898 ACCCTGCGCGCGAGCAAGGCGCAACCCAGAGGTGAAGAATGTGATGACCGACACCTGCTG 957  
QY 961 GTGCAGAAAGCCCAACCCCGCACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCAGC 1020  
DB 958 GTGCAGAAAGCCCAACCCCGCACTGCAAGACCATCTGCGGCTCTGCGGCCCGCGCAGC 1017  
QY 1021 CTGAGAGATGATGACCCGCTGCGAGGCGTGGCGCGCCAGCCCAAGGCCGCGTG 1080  
DB 1018 CTGAGAGATGATGACCCGCTGCGAGGCGTGGCGCGCCAGCCCAAGGCCGCGTG 1077  
QY 1081 CTGCGCGAGCGATGAGCGAGCCCAACAC---AGCGTATGATGACAGAGCAACTTC 1137  
DB 1078 CTGCGCGAGCGSATGAGCGAGCCCAACAGCGTGAACATCATGATGACAGAGCAACTTC 1137  
QY 1138 AAGGCGCCCGCGCATCGTCAAGTGTCTCACTGCGGCAAGAGAGGCGCACATCGCCGC 1197  
DB 1138 AAGGCGCCCGCGCAACGTSAGTGTCTCACTGCGGCAAGAGAGGCGCACATCGCCAG 1197  
QY 1198 AACTGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGTGCGCAAGAGAGGCGCACAGATG 1257  
DB 1198 AACTGCGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGTGCGCAAGAGAGGCGCACAGATG 1257  
QY 1258 AAGACTGCAACCGAGCGCGCAAGGCGCACTTCTGGGCAAGATCTGCGCCAGCAAGGGC 1317  
DB 1258 AAGACTGCAACCGAGCGCGCAAGGCGCACTTCTGGGCAAGATCTGCGCCAGCAAGGGC 1317  
QY 1318 CGCCCCGCACTTCTCTGAGAGCGG-----CCCCGAGCCCAAGCGCC 1359  
DB 1318 CGCCCCGCACTTCTCTGAGAGCGGAGCGAGCGCGCGCCCGCCCAAGCGCGCC 1377  
QY 1360 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCCAGAGAGAGAGCAAG 1419  
DB 1378 CCCCCCGCGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCCAGAGAGAGAGCAAG 1437  
QY 1420 G-----ACCGGAGACCTTGACCAAGCTGGAAGCTTGTGGCAAGACCC 1467  
DB 1438 GACCGGAGGCGCTTACCGGAGCGCCCTGACCGCCCTGCGGAGCGCGGCC 1497  
QY 1468 CTGAGCCAGTAA 1479  
DB 1498 CTGAGCCAGTAA 1509

RESULT 11  
AX455905 1509 bp DNA linear PAT 06-JUL-2002  
LOCUS AX455905  
DEFINITION Sequence 21 from Patent WO0204493.  
ACCESSION AX455905  
VERSION AX455905.1 GI:21714897  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
synthetic construct  
other sequences; artificial sequences.  
REFERENCE  
1 zur Megeide,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,B.  
AUTHORS Polynucleotides encoding antigenic hiv type c polypeptides,  
TITLE polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 21 17-JAN-2002;  
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
FEATURES  
source 1. .1509  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="synthetic Gag coding sequence of HIV strain  
AF110967"

ORIGIN

Query Match		85.2%;	Score 1260.8;	DB 6;	Length 1509;
Best Local Similarity		91.9%;	Pred. No. 3.2e-127;		
Matches 1389;		Conservative 0;	Mismatches 87;	Indels 36;	Gaps 4;
QY	1	ATGGGGCGCCGCGCCAGCATCTCTGGCGGGCGGCGCAAGCTGAGCCCTGGAGCGCATCCGC	60		
DB	1	ATGGGGCGCCGCGCCAGCATCTCTGGCGGGCGGCGAGAGCTGAGCAAGTGGAGAAATCCGC	60		
QY	61	CTGGCGCCCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGCCAGCGCGAG	120		
DB	61	CTGGCGCCCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGCCAGCGCGAG	120		
QY	121	CTGGAAGATTGCTCTGAACCCCGGCTGTGAGAGCCAGCGAGGGCTGCAAGCATC	180		
DB	121	CTGGAAGCTTCTGCTCTGAACCCCGGCTGTGAGAGCCGCGAGGGCTGCAAGCATC	180		
QY	181	ATCCGCGAGCTGACCCCGCTGACAGCCGCGAGGAGCTGAAGAGCTGTTCAC	240		
DB	181	ATGAAGCAGCTGACCCCGCTGACAGCCGCGAGGAGCTGCGCAGCTGTACAC	240		
QY	241	ACCGTGCCACCTGTACTGCTGTCAGAGAAAGTGAAGTCCGCGACACCAAGAGGCC	300		
DB	241	ACCGTGCCACCTGTACTGCTGTCAGCCGCGCATCGAGGTGCGGACACCAAGAGGCC	300		
QY	301	CTGGAACAAGATCGAGAGAGCAACAAGTGCAGACAGAAATCCAGAGCGCGAGCC	360		
DB	301	CTGGAACAAGATCGAGAGAGCAACAAGAGCCAGAGAACCCAGAGCGCGAGAG	360		
QY	361	GCCGACAAGGCAAGTGAAGCCAGAACTACCCCATCTGTGAGAACTTGAGGGCCAGATG	420		
DB	361	GCCGAC---GGCAAGTGAAGCCAGAACTACCCCATCTGTGAGAACTTGAGGGCCAGATG	417		
QY	421	GTGACACGAGCCATCAGCCCGCGACCTGAAGCCTGGTGAAGTGTGAGAGAG	480		
DB	418	GTGACACGAGCCATCAGCCCGCGACCTGAAGCCTGGTGAAGTGTGAGAGAG	477		
QY	481	GCCTTCAGCCCGGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCACCCCCAG	540		
DB	478	GCCTTCAGCCCGGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCACCCCCAG	537		
QY	541	GACCTGAACAGATGTTGAACACCGTGGCGGCGCACAGCGCGCATGAGATGTGAAG	600		
DB	538	GACCTGAACACCATGTGAACACCGTGGCGGCGCACAGCGCGCATGAGATGTGAAG	597		
QY	601	GACACCATCAAGAGAGGCGCGGAGTGGACCGCGTGACCCCGTGACGCGCGCCCC	660		
DB	598	GACACCATCAACAGAGAGGCGCGGAGTGGACCGCGTGACCCCGTGACGCGCGCCCC	657		
QY	661	ATCGCCCCCGGCGAGATGCGCGAGCCCGCGGCGAGCATCGCGCGCACCAAGCAC	720		
DB	658	GTGGCCCCCGGCGAGATGCGCGAGCCCGCGGCGAGCATCGCGCGCACCAAGCAC	717		
QY	721	CTGCAAGAGAGATCGCTGTGATCAAGCAACCCCGCATCCCGTGGGCGCATCTAC	780		
DB	718	CTGCAAGAGAGATCGCTGTGATCAAGCAACCCCGCATCCCGTGGGCGCATCTAC	777		
QY	781	AAGCGGTGATCATCTGGGCTGTGAACAAGATCGTGGCGGATGTACAGCCCGTGAATC	840		
DB	778	AAGCGGTGATCATCTGGGCTGTGAACAAGATCGTGGCGGATGTACAGCCCGTGAATC	837		
QY	841	CTGGAATCAAGAGAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAGCCGTTCTCAAG	900		
DB	838	CTGGAATCAAGAGAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAGCCGTTCTCAAG	897		
QY	901	ACCCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	960		
DB	898	ACCCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	957		
QY	961	GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1020		
DB	958	GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1017		
QY	1021	CTGGAAGAGATGATGACCGCTGCAAGGGCGTGGGGCGGCCCAAGCAAGGCCCGCTG	1080		

DB	1018	CTGAGAGAGATGATGACCGCTGCGCAGGCGGTGGCGGGCCCGGCCACAAAGGCCGCTG	1077		
QY	1081	CTGCGCGAGCGCATGAGCCAGGCCAACACC---AGCGTATGATGACAGAAAGCAATTC	1137		
DB	1078	CTGCGCGAGCGCATGAGCCAGGCCAACAGCTGAACATCATGATGACAGAAAGCAATTC	1137		
QY	1138	AAGGCCCCCGCGCATGTGTAGTCTTCACTGCGGCAAGAGAGGCGCATCGCCGC	1197		
DB	1138	AAGGCCCCCGCGCAACGTGAAGTCTTCACTGCGGCAAGAGAGGCGCATCGCCAAAG	1197		
QY	1198	AACTGCGCGCGCCCGCAAGAGGCTGTGGAAGTGCAGGCAAGAGGCGCACAGATG	1257		
DB	1198	AACTGCGCGCGCCCGCAAGAGGCTGTGGAAGTGCAGGCAAGAGGCGCACAGATG	1257		
QY	1258	AAGACTGACCGAGCGCCAGGCCAACTTCTGGGCAAGATGCCCCAGCCCAAGGGC	1317		
DB	1258	AAGACTGACCGAGCGCCAGGCCAACTTCTGGGCAAGATGCCCCAGCCCAAGGGC	1317		
QY	1318	CGCCCCGCAACTTCTGACAGCGG-----CCCCGAGCCCAAGGCC	1359		
DB	1318	CGCCCCGCAACTTCTGACAGACCGAGAGCGCCGCGCCCAACGTGCCACCGCC	1377		
QY	1360	CCCCCGCGGAGAGCTTCCGCTTGAAGAGACACCCCGGCGCAAGAGAGAGCAAG	1419		
DB	1378	CCCCCGCGGAGAGCTTCCGCTTGAAGAGACACCCCGGCGCAAGAGAGAGCAAG	1437		
QY	1420	G-----ACCGGAGACCTTGACCAAGCCTGAAGAGCTGTTGCGCAAGACCCC	1467		
DB	1438	GACCGGAGCCCTTACCGGAGCGCCCTGACCGCGCTGCGAGCTGTTGCGAGCGGCCCC	1497		
QY	1468	CTGAGCCAGTAA 1479			
DB	1498	CTGAGCCAGTAA 1509			

RESULT 12  
AX427930 9166 bp DNA linear PAT 20-JUN-2002  
LOCUS AX427930  
DEFINITION Sequence 168 from Patent WO0232943.  
ACCESSION AX427930  
VERSION AX427930.1 GI:21538017  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
synthetic construct  
other sequences; artificial sequences.  
REFERENCE  
1 Huang,Y. and Nabel,G.J.  
AUTHORS  
TITLE  
Modifications of hiv env, gag, and pol enhance immunogenicity for  
JOURNAL  
Genetic Immunization  
Patent: WO 0232943-A 168 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
FEATURES  
source  
1. 9166  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

Query Match		82.6%;	Score 1221.2;	DB 6;	Length 9166;
Best Local Similarity		90.1%;	Pred. No. 3.5e-123;		
Matches 1346;		Conservative 0;	Mismatches 133;	Indels 15;	Gaps 3;
QY	1	ATGGGGCGCCGCGCCAGCATCTCTGGCGGGCGGCAAGCTGACGCTGGAGCGCATCCGC	60		
DB	1880	ATGGGGCGCCGCGCCAGCATCTCTGGCGGGCGGCAAGCTGACGCTGGAGCGCATCAAG	1939		
QY	61	CTGGCGCCCGCGGCAAGAGTGTACATGATGAAGACCTGTGTGGGCGAGCGCGAG	120		
DB	1940	CTGAAGCCCGGCGCAAGAGCACTACATGATGAAGACCTGTGTGGGCGAGCGCGAG	1999		
QY	121	CTGGAAGAATTCCGCTGAACCCCGGCTGTGAGAGACGAGAGGCTGCAAGCATC	180		

Db 2000 CTGGAGCGCTTGCCTTGAGACCCCGGCTGTGAGAGACCAAGGGGCTGCAAGCATC 2059  
 QY 181 ATCCGCCAGTGCACCCCGCTGACAGCCGGCAGAGAGCTGAAGAGCTGTTCAAC 240  
 Db 2060 ATGAAGCAGCTGACCCCGCTGACAGCCGGCACAAGAGCTGATCAGCTTGACAAC 2119  
 QY 241 ACCGTGGCACCTGTACTGCGTGCAGAGAAATCGAGTCCGCGACACCAAGAGGCC 300  
 Db 2120 ACCGTGGCACCTGTACTGCGTGCAGAGAAATCGAGTCCGCGACACCAAGAGGCC 2179  
 QY 301 CTGGAACAAGATCGAGAGAGAGACAAGTGCAGACGAAGATCCAGAGCGCGAGGCC 360  
 Db 2180 CTGGAACAAGATCGAGAGAGAGACAAGAGCCAGACGAAGACCAAGAGCGCGAGGCC 2239  
 QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCTGACAGAACTGACAGGCGCATG 420  
 Db 2240 GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCTGACAGAACTGACAGGCGCATG 2299  
 QY 421 GTGCACCAAGGCGCATCAGCCCCCGACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480  
 Db 2300 GTGCACCAAGGCGCATCAGCCCCCGACCCCTGAACGCTGGGTGAAGGTGAGAGAAG 2359  
 QY 481 GCCTTCAGCCCCGAGGTGATCCCATATGTTCAACGCGCTGAGCGAGGGCGCCACCCCGCAG 540  
 Db 2360 GCCTTCAGCCCCGAGGTGATCCCATATGTTCAACGCGCTGAGCGAGGGCGCCACCCCGCAG 2419  
 QY 541 GACCTGAACAAGATGTTGAACAACCGTGGCGGCCACCAAGCGCCATGACAGATGCTGAAG 600  
 Db 2420 GACCTGAACAAGATGTTGAACAACCGTGGCGGCCACCAAGCGCCATGACAGATGCTGAAG 2479  
 QY 601 GACACCATCAACGAGAGAGAGCGCCCGAGTGGGACCGCGCTGCACCCCGTGCACGCGCGGCC 660  
 Db 2480 GACACCATCAACGAGAGAGAGCGCCCGAGTGGGACCGCGCTGCACCCCGTGCACGCGCGGCC 2539  
 QY 661 ATCGCCCCCGGCAAGATGCGCGAGCCCCCGCGGAGCGACATCGCGCGGCAACCAAGCACC 720  
 Db 2540 ATCGCCCCCGGCAAGATGCGCGAGCCCCCGCGGAGCGACATCGCGCGGCAACCAAGCACC 2599  
 QY 721 CTGCAGAGAGATCGCGCTGATGACCAAGCAACCCCGCATCCCGCTGGGCGCATCTAC 780  
 Db 2600 CTGCAGAGAGATCACTGATGACCAACCAACCCCGCATCCCGCTGGGCGCATCTAC 2659  
 QY 781 AAGCGGTGATCATCTGCGGCTGACAAAGATGTCGGATGTAACAGCCCGTGAAGCATC 840  
 Db 2660 AAGCGGTGATCATCTGCGGCTGACAAAGATGTCGGATGTAACAGCCCGTGAAGCATC 2719  
 QY 841 CTGCATCAAGCAGGAGGCGCCCAAGAGCCCTTCGCGACTACGTGGAACCGCTTCTCAAG 900  
 Db 2720 CTGCATCAAGCAGGAGGCGCCCAAGAGCCCTTCGCGACTACGTGGAACCGCTTCTCAAG 2779  
 QY 901 ACCCTGCGCGCGGAGCAGAGCAACCAAGAGGTGAAGAACTGATGACCGACACCTGCTG 960  
 Db 2780 ACCCTGCGCGCGGAGCAGAGCAACCAAGAGGTGAAGAACTGATGACCGACACCTGCTG 2839  
 QY 961 GTGCAGAAACGCAACCCCGACTGCAAGAACCATCTGCGGCTCTCGGCCCCGGGCGCCAGC 1020  
 Db 2840 GTGCAGAAACGCAACCCCGACTGCAAGAACCATCTGCGGCTCTCGGCCCCGGGCGCCAGC 2899  
 QY 1021 CTGAGAGAGATGATGACCGCTGCGAGGGCGGTGGGCGGCGCCCAAGAGCGCGCTG 1080  
 Db 2900 CTGAGAGAGATGATGACCGCTGCGAGGGCGGTGGGCGGCGCCCAAGAGCGCGCTG 2959  
 QY 1081 CTGCGCGAGGCGATGAGCCAGG--CCAACACAGCGGTGATGATGCAAGAGCAACTTC 1137  
 Db 2960 CTGCGCGAGGCGATGAGCAAGGTGAACAACAACATCATGATGCAAGCGCAAGCACTGC 3019  
 QY 1138 AAGGGCCCCCGGCGCATCTGTAAGTCTTCAACTGCGGCAAGAGAGGGCCACATCGCCCGC 1197  
 Db 3020 AAGGGCCCCCGGCGCATCTGTAAGTCTTCAACTGCGGCAAGAGAGGGCCACATCGCCCGC 3079  
 QY 1198 AACTGCGCGCGCGCGCAAGAGGGGTGCTGGAAGTGGCGCAAGAGGGCCACCATG 1257

Db 3080 AACTGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCATG 3139  
 QY 1258 AAGACTGACACCGAGCGCGCCCAACTTCTCTGGGCAAGATCTGGCCAGCCCAAGGGC 1317  
 Db 3140 AAGACTGACACCGAGCGCGCCCAACTTCTCTGGGCAAGATCTGGCCAGCCCAAGGGC 3199  
 QY 1318 CGCCCCGCAACTTCTCTGACAGAGCGCGCCGAGCCCAAGCGCGCGCGCGAGAGCTTC 1377  
 Db 3200 AGCCAGGGAATTTCTTTCAGAGCAGACCAAGCCCAAGAGCGCGCGCGAGAGAGCTTC 3259  
 QY 1378 CGCTTCAGAGG-----GACCAACCGCGCGCGCAAGAGCAGAG-----AGCAAGAGCCGC 1425  
 Db 3260 AGGTTGGGGAAGAGACAACAACCTCCCTCTCAGAAAGCAGAGCGCGATAGACAAGAACTG 3319  
 QY 1426 GAGACCCCTGACCAAGCTGAAGAGCGCTGTCGCAAGCAGACCCCTGAGCCAGTAA 1479  
 Db 3320 TATCCTTAGCTTCCCTCAGATCACTCTTGGCAGCAGACCCCTGTCACAACTAA 3373

RESULT 13  
 AX427936 9788 bp DNA linear PAT 20-JUN-2002  
 LOCUS AX427936  
 DEFINITION Sequence 174 from Patent WO0232943.  
 ACCESSION AX427936  
 VERSION AX427936.1 GI:21538023  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 synthetic construct  
 other sequences; artificial sequences.  
 REFERENCE  
 1 Huang, Y. and Nabel, G.J.  
 Modifications of hiv env, gag, and pol enhance immunogenicity for  
 genetic immunization  
 Patent: WO 0232943-A 174 25-APR-2002;  
 GOVERNMENT OF THE UNITED STATES (US)  
 FEATURES  
 source  
 1. 9788  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="plasmid pVR1012x/s containing HIV genes"

ORIGIN  
 Query Match 81.5%; Score 1206; DB 6; Length 9788;  
 Best Local Similarity 90.0%; Pred. No. 1.5e-121;  
 Matches 1344; Conservative 0; Mismatches 130; Indels 20; Gaps 4;  
 QY 1 ATGGGCGCGCGCGCGCAGCATCTCGCGCGCGCAGCTGGAACGCTTGGAGCGCATCCGC 60  
 Db 1880 ATGGGCGCGCGCGCGCAGCATCTCGCGCGCGCAGCTGGAACGCTTGGAGCGCATCAAG 1939  
 QY 61 CTGCGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGGCCAGCGCGAG 120  
 Db 1940 CTGAAGCGCGCGCGCGCAAGAGCACTACATGATGAAGCACTGTGTGGGCCAGCGCGAG 1999  
 QY 121 CTGGAAGAGTTGCGCCCTGAACCCCGGCTGTGAGAGCAAGCGGCTGCAAGCATC 180  
 Db 2000 CTGAAGCGCTTCGCCCTGAGACCCCGGCTGTGAGAGCAAGCGGCTGCAAGCATC 2059  
 QY 181 ATCGCGCAGCTGCACCCCGCGCTGCAAGACCGGCAAGGAGCTGAAGAGCTGTCAAC 240  
 Db 2060 ATGAAGCAGCTGCAGCCCGCGCTGCAAGACCGGCAAGGAGCTGATCAGCTGCAAC 2119  
 QY 241 ACCGTGGCCACCTGTACTGCGTGCAGAGAGATCGAGTCCGCGACACCAAGAGGCC 300  
 Db 2120 ACCGTGGCCACCTGTACTGCGTGCAGAGAGATCGAGTCCGCGACACCAAGAGGCC 2179  
 QY 301 CTGGAACAAGTTCGAGAGAGAGACAAGTGCAGCAAGATCCAGAGAGCTCAGAGAGGCC 360  
 Db 2180 CTGGAACAAGTTCAGAGAGAGAGACAAGTGCAGCAAGATCCAGAGAGCTCAGAGAGGCC 2239  
 QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCTGACAGAACTGACAGGCGCATG 420

Db 2240 GCCGACAAGGGCAAGGTGAGCCAGAACTAACCCTATCGTGCAGAACTCTGCAGGGCCAGATG 2299

QY 421 GTGCACCAAGGCCATCAGCCCCCGCACCTTGAAACGCTGGGTGAAGGTGATCGAGAGAA 480

Db 2300 GTGCACCAAGGCCATCAGCGCCCGCACCTTGAAACGCTGGGTGAAGGTGATCGAGAGAA 2359

QY 481 GCCTTCAGCCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCCACCCCCAG 540

Db 2360 GCCTTCAGCCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCCACCCCCAG 2419

QY 541 GACCTGAACACGATGTTGAACACCCGTGGGGCGGCCACAGGGCCCATGCGATGCTGAAG 600

Db 2420 GACCTGAACACGATGTTGAACACCCGTGGGGCGGCCACAGGGCCCATGCGATGCTGAAG 2479

QY 601 GACACCATCAACGAGAGAGGGCCCGGAGTGGGACCCGCTGCACCCCGTGCACGCGCGCCC 660

Db 2480 GACACCATCAACGAGAGAGGGCCCGGAGTGGGACCCGCTGCACCCCGTGCACGCGCGCCC 2539

QY 661 ATGCCCCCGGCGAGATGCGCGAGCGCCGCGGAGCGACATCGCCGCGACCAACAGCAC 720

Db 2540 GTGCCCCCGGCGAGATGCGCGAGCGCCGCGGAGCGACATCGCCGCGACCAACAGCAC 2599

QY 721 CTGCAGGAGCAGATCGCTGTGATGACCAACCCCCCATCCCCCGTGGGCGACATCTAC 780

Db 2600 CTGCAGGAGCAGATCACTGTGATGACCAACCCCCCGTGGGCGACATCTAC 2659

QY 781 AAGCGGTGATCATCTCTGGGCTGTGAACAAGATCGTGGGATGTACAGCCCCGTGACATC 840

Db 2660 AAGCGGTGATCATCTCTGGGCTGTGAACAAGATCGTGGGATGTACAGCCCCGTGACATC 2719

QY 841 CTGCACATCAAGCAGGGGCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 900

Db 2720 CTGCACATCAAGCAGGGGCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 2779

QY 901 ACCCTGCGCGCGCAGAGCAGACCAAGAGGTGAAGAACTGATGACCGACACCTGCTG 960

Db 2780 ACCCTGCGCGCGCAGAGCAGACCAAGAGGTGAAGAACTGATGACCGACACCTGCTG 2839

QY 961 GTGCAGAACGCAACCCCGACTGCAAGCCATCTGCGCGCTCTCGGCCCGCGCGCAGC 1020

Db 2840 GTGCAGAACGCAACCCCGACTGCAAGCCATCTGCGCGCTCTCGGCCCGCGCGCAGC 2899

QY 1021 CTGCAGAGAGATGATGACCGCTGCGAGGGCGGTGGCGGCCCAAGGCGCGGTG 1080

Db 2900 CTGCAGAGAGATGATGACCGCTGCGAGGGCGGTGGCGGCCCAAGGCGCGGTG 2959

QY 1081 CTGCGCGAGGCGATGAGCCAGG--CCAACACCAAGCGTGTATGTGCAAGAACAACTTC 1137

Db 2960 CTGCGCGAGGCGCATGAGCAAGGTGAACAACACCAACATCATGATGACGCGACGAACTGC 3019

QY 1138 AAGGGCCCCCGGCGCATCTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGC 1197

Db 3020 AAGGGCCCCCGGCGCATCTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGC 3079

QY 1198 AACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGGGCCACAGATG 1257

Db 3080 AACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGGGCCACAGATG 3139

QY 1258 AAGGACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCCAAGCCAAAGGGC 1317

Db 3140 AAGGACTGCACCGAGCGCAAGGCTTA-----TAGGGAAGATCTGGCTTCCCAAGGGA 3194

QY 1318 CGCCCCGGCACTTCTGCAAGAGCGCCCCCGAGCCACCGCCCCCGCGAGAGCTTC 1377

Db 3195 AGGCCAAGGGAATTTCTTCAAGAGCAACAGAGCCCAAGCCCAAGAGAGAGCTTC 3254

QY 1378 CGCTTCGAGGA-----GACCACCCCCGGCGCAAGAGCAGAG-----AGCAAGAGCCGC 1425

Db 3255 AGCTTTGGGGAAGAGCAACAATCTCTCTCAAGAGCAGAGCCGATAGACAAAGGAATG 3314

QY 1426 GAGACCCCTGACCAAGCCTGAAGAGCCTGTTCGCAACGACCCCTGAGCCAGTAA 1479

Db 3315 TATCTTTAGCTTCCCTCAGATCACTTTTGGCAGGCAACCCCTCGTCAATATA 3368

RESULT 14

AX149648

LOCUS AX149648 4288 bp DNA linear PAT 08-JUN-2001

DEFINITION Sequence 2 from Patent WO0136614.

ACCESSION AX149648

VERSION AX149648.1 GI:14348047

KEYWORDS

SOURCE

ORGANISM

Human immunodeficiency virus

Human immunodeficiency virus

Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate

lentivirus group.

REFERENCE

1

Shao,Y., Wagner,R., Wolf,H. and Graf,M.

The genome of the hiv-1 inter-subtype (c/b') and use thereof

Patent: WO 0136614-A 2 25-MAY-2001;

Geneart GMBH Gesellschaft fuer angewandte Biotechnologie (DE) ;

Shao, Yiming (CN)

Location/Qualifiers

1. .4288

/organism="Human immunodeficiency virus"

/mol\_type="unassigned DNA"

/db\_xref="taxon:12721"

FEATURES

source

ORIGIN

Query Match 81.5%; Score 1205.4; DB 6; Length 4288;

Best Local Similarity 89.5%; Pred. No. 2.2e-121;

Matches 1335; Conservative 0; Mismatches 141; Indels 15; Gaps 3;

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGCGGCAAGCTGACCGCTGGAGCGCATCCG 60

Db 13 ATGGGCGCGAGGCGCAGCATCTGAGGGGCGGCAAGCTGACCAAGTGGAGAGATCAGG 72

QY 61 CTGCGCCCCGGCGCAAGAGTGTATCATGATGAGCAACCTGTGTGGCCAGCCGCGAG 120

Db 73 CTGAGGCGCGCGCAAGAGCACTATGCTGAAGCACCTGTGTGGCCAGCAGAGAG 132

QY 121 CTGCAGAGATTGCGCTTGAACCCCGGCTGTGAGAACAGCGAGGGCTGCAAGCAGATC 180

Db 133 CTGCAGAGATTGCGCTTGAACCCCGGCTGTGAGAACAGCGAGGGCTGCAAGCAGATC 192

QY 181 ATCCGCAAGCTGCAACCCCGCTGCAAGCCGCGAGCGAGAGCTGAAGAGCTGTTCAAC 240

Db 193 ATGAAGCAGCTGCAAGCGCTGCAAGACCGGCAACGAGAGCTGAAGAGCTGTTCAAC 252

QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAAAGATCGAGTCCGCGACCAAGAGAGCC 300

Db 253 ACCGTGGCCACCTCTACTGCGTGCACACCGAATCGAGTGAAGGACACAGGGAGGCC 312

QY 301 CTGCACAAGATCGAGAGAGAGCAACAAGTGCCAGCAGAAAGATCCAGAGCGCGAGGCC 360

Db 313 CTGCACAAGATCGAGAGAGAGCAACAAGATCCAGCAGAAAGACCAAGAGCGCAAGAG 372

QY 361 GCCGCAAGGGGCAAGGTGAGCGCAACTACCCCATCTGCAAGAACTGCAAGGGCCAGATG 420

Db 373 GCCGAC--GGCAAGGTGAGCGCAACTACCCCATCTGCAAGAACTGCAAGGGCCAGATG 429

QY 421 GTGCACCAAGGCCATCAGCCCCCGCACCTTGAACGCTGGGTGAAGGTGATGAGGAGAG 480

Db 430 GTGCACCAAGGCCATCAGCCCCCGCACCTTGAATGCATGGGTGAAGGTGATGAGGAGAG 489

QY 481 GCCTTCAGCCCCCGAGGTGATCCCATGTTCAACGCCCCGTGAGCGAGGGGCCACCCCCAG 540

Db 490 GCCTTCAGCCCCCGAGGTGATCCCATGTTCAAGCGCCCTGAGCGAGGGGCCACCCCCAG 549

QY 541 GACCTGAACAAGATGTTGAACAACCGTGGGCGGCCACCAAGCGCCCATGAGATGCTGAAG 600

Db 550 GACCTGAACAACATGCTGAACAACCGTGGGCGGCCACCAAGCGCCCATGAGATCTGAAG 609

QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCCC 660

Db 610 GACACCATCAACGAGAGAGCGCGCGAGTGGGACGAGCTGCACCCCGTGCACGCGCGCCCC 669

QY	661	ATCGCCCCCGCCAGATGCGCGAGCCCCCGGCGAGCGACATCGCGGGCAACCAACGACACC	720
Db	670	ATCGCCCCCGCCAGATGAGGGAGCCCCAGGGGCAAGCATCGCTGGCAACCAACGACAAAC	729
QY	721	CTGCAGGAGCAGATCGCTTGATGACCAAGCAACCCCCCATCCCCGTGGGCGCATCTAC	780
Db	730	CTGCAGGAGCAGATCGCTTGATGACCAAGCAACCCCCCGTGCCGTGGGCGCATCTAC	789
QY	781	AAGCGGTGATCATCTGGGCGTTAAACAAGATCGTGCGGATGTACAGCCCCCGTAGCATC	840
Db	790	AAGAGGTGATCATCTGGGTTTAAACAAGATCGTAGGATGTACAGCCCCCAACGACATC	849
QY	841	CTGCACATCAAGAGGGGGCCCAAGAGAGCCCTTCCCGCACTACCGTGAGCCGCTTCTCAAG	900
Db	850	CTGCACATCAAGAGGGGGCCCAAGAGAGCCCTTCAAGGACTACGTGACAGGTTCTTCAAG	909
QY	901	ACCGTCGCGCGCGAGCAGAGCAACCAAGAGGTGAAGAACTGATGACCGACACCTGCTG	960
Db	910	ACCGTGAAGGCGGAGCAGGCCAACCAAGGCGTGAAGAACTGATGACCGACACCTGCTG	969
QY	961	GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGGCCCCGGGCGCAGC	1020
Db	970	GTGCAGAACGCCAACCCCGACTGCAAGACCATCTGAAGGCGCTGCGGGCCCCGGGCGCAGC	1029
QY	1021	CTGGAGGATGATGACCGCGCTGCCAGGGCGTGGGCGGGCCCCAGCCACAAGGCCGCGTG	1080
Db	1030	ATCGAGGAGATGATGACCGCGCTGCCAGGGCGTGGGCGGGCCCCAGCCACAAGGCCAAGTGTG	1089
QY	1081	CTGGCGGAGGCGATGAGCCAGGCGCAACACCAAGCGTATGATGCAAGAGCAACTTCAAG	1140
Db	1090	CTGGCGGAGGCGCATGAGCCAGACCAACAGCGCATCTGATGCAAGAGCAACTTCAAG	1149
QY	1141	GGCCCCCGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCCGCAAC	1200
Db	1150	GGCAGCAAGAGATCGTGAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCAGGAAC	1209
QY	1201	TGCGCGCCCCCGCCCAAGAGGCTGTGAAGTGCAGGCAAGGAGGGCCACGATGAAG	1260
Db	1210	TGCAAGGGCCCCCGAAGAAAGAGGCTGTGAAGTGCAGGCAAGGAGGGCCACGATGAAG	1269
QY	1261	GACTGCACCGAGCGCCAGGCCAACTTCTTGGGCAAGATCTGGCCAGCCACAAGGGCGCG	1320
Db	1270	GACTGCACCGAGGCGCAGGCCAACTTCTTGGGCAAGATCTGGCCAGCCACAAGGGCGCG	1329
QY	1321	CCCGGCAACTTCTGCAAGAGCGCGCCCGAGCCACCGCCCCCGCGCGAGAGCTTCCGC	1380
Db	1330	CCCGGCAACTTCTGCAAGAACAGGCGCCGAGCCCAACGCCCCCGCGAGAGAGCTTCAAG	1389
QY	1381	TTCGAGGAG-----ACCACCCCGGGCCAGAAAGCAGGAG-----AGCAAGGACCGCGAG	1428
Db	1390	TTCGAGGAGGAGACCAACACCCCGCAGCGAAGCAGGAGCCCATGCAACAAGAGCTGTAC	1449
QY	1429	ACCCTGACCAAGCCTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAGTAA	1479
Db	1450	CCCTGACCAAGCCTGAAGAGCCTGTTCGGCAACGACCCCAAGCAGCAAGAA	1500

RESULT 15	AF201927	LOCUS	1509 bp	DNA	linear	SVN	16-MAR-2000
DEFINITION	Synthetic construct gag protein gene, complete cds.						
ACCESSION	AF201927						
VERSION	AF201927.1						
KEYWORDS	GI:7248702						
SOURCE	synthetic construct						
ORGANISM	synthetic construct						
REFERENCE	1 (bases 1 to 1509)						
AUTHORS	zur Megede,J., Chen,M.C., Doe,B., Schaefer,M., Greer,C.E., Selby,M., Otten,G.R. and Barnett,S.W.						
TITLE	Increased expression and immunogenicity of sequence-modified human immunodeficiency virus type 1 gag gene						

JOURNAL	J. Virol. 74 (6), 2628-2635 (2000)	
MEDLINE	20148954	
PUBMED	10684277	
REFERENCE	2 (bases 1 to 1509)	
AUTHORS	zur Megede,J. and Barnett,S.W.	
TITLE	Direct Submission	
JOURNAL	Submitted (04-NOV-1999) Vaccines, Chiron Corporation, 4560 Horton, Emeryville, CA 94608, USA	
FEATURES	Location/Qualifiers	
source	1..1509	
	/organism="synthetic construct"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:32630"	
	1..1509	
	/note="HIV-1SF2 p55gag"	
	/codon_start=1	
	/transl_table=11	
	/product="gag protein"	
	/protein_id="AAFA3628.1"	
	/db_xref="GI:7248703"	
	/translation="MGARASVLSGSELDKWEKIRLRPGKKYKLIKHIWASRELER AVNPGLETSEGRQILGOLPSLPTGSEBELRSYNTVATLYVHORIDVKDTKEALE KIEEONKSKKKAQAAAAAGTENSQVSQNPVYNLQGMVHQAISPTLNAYKV VEEKAQSPVIMFSALESGATPDNLMTNVGHAQAMOMLKETINEEAEWDRVH PVHAGPIAPGMREPRGSDIAGTSTLQEQIGMTNPNPIPVGEIYKRWIILGLNKIV RMVSPSILDIRQKPEPRDYVRFYKTLRAEQASQDVKNMTETLLVNANPCKT ILKALGPATLIEEMMTACQGVGPGHKARVLAEANSQVTNPATIMQGNFRNQKTV KFCNGKEGHTANCRAPRKKGCMGREGHOMQCTERGHOMFLGKIMPSYKGRGNF LQSRPEPTAPPEESFRFGEKTPSQKQEPIDKEIYPLTSLRSLFGNDPSSQ"	
ORIGIN		
Query Match	81.0%; Score 1197.8; DB 12; Length 1509;	
Best Local Similarity	89.3%; Pred. No. 2e-120;	
Matches 1347; Conservative	0; Mismatches 132; Indels 30; Gaps 4;	
QY	1 ATGGGCGCCCGCGCCAGCATCTGCGCGCGCAAGCTGACGCTGGAGCGCATCCGC 60       DB	1 ATGGGCGCCCGCGCCAGCGTGTGAGCGCGCGAGCTGACAAAGTGGAGAATCCGC 60 
QY	61 CTGCGCCCCGCGCGCAAGAGTGCTACATGATGAAGCACCTGCTGGCGCGAG 120       DB	61 CTGCGCCCCGCGCGCAAGAGTAGTACAAGCTGAAGCACATGCTGTGGCGCGAG 120 
QY	121 CTGGAAGAAGTTCGCTGAACCCCGGCGCTGTGAGAGACGAGGAGGCTGCAAGCATC 180       DB	121 CTGGAAGCGCTTCGCGTGAACCCCGGCGCTGTGAGAGACGAGGAGGCTGCGCAGATC 180 
QY	181 ATCCGCCAGCTGCACCCCGCCTGCAGACCGGCAAGGAGAGCTGAAGAGCTGTCAAC 240       DB	181 CTGGGCGCAGCTGCAGCCAGCCTGCAGACCGGCAAGGAGAGCTGCGCAGCTGTACAAC 240 
QY	241 ACCGTGCCACCTGTACTGCGTGCACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300       DB	241 ACCGTGCCACCTGTACTGCGTGCACGAGCGCATCGACGTCAAGAGACACCAAGAGGCC 300 
QY	301 CTGGAACAAGATCGAGAGGAGCAGAACAAGTSCAGCAGAAAGATCCAGCAGGCGGAGCC 360       DB	301 CTGGAAGAAGATCGAGAGGAGCAGAACAAGTCCAAAGAAAGGCCAGCAGGCGCGGCC 360 
QY	361 GCCG-----ACAAAGGCAAGGTGAGCCAGAACTACCCATCGTGAGAAACCTG 408       DB	361 GCCGCGGCAACCGGCAACAGCAGCAGGTGAGCCAGAACTACCCATCGTGAGAAACCTG 420 
QY	409 CAGGCGCAGATGCTGACCAAGGCGCATCAGCCCCGCGACCTGAAGCGCTGGTGAAGGTG 468       DB	421 CAGGCGCAGATGCTGACCAAGGCGCATCAGCCCCGCGACCTGAAGCGCTGGTGAAGGTG 480 
QY	469 ATCGAGAGAGAGGCTTTCAGCCCCGAGGTGATCCCATGTTTCAACGCGCTGAGCGAGGGC 528       DB	481 GTGAGAGAGAGGCTTTCAGCCCCGAGGTGATCCCATGTTTCAAGCGCTGAGCGAGGGC 540 
QY	529 GCCACCCCGCAGGACTGAACAAGATGTTGAACAACGTGGCGGCGCACGAGCGCCCATG 588 	

```
Db      541 GCCACCCCCCAGGACCTGGAACACGATGTTGAACACCGTGGGGCCCAACAGGCCGCATG 600
QY      589 CAGATGCTGAAAGACACATCAACGAGAGGCCGCCGAGTGGGACCGCGTGCAACCCCGTG 648
Db      601 CAGATGCTGAAAGAGACCATCAACGAGAGGCCGCCGAGTGGGACCGCGTGCAACCCCGTG 660
QY      649 CACGCCGGGCCCCCATCGCCCCCGGGCCAGATGCGCGAGCGCCCGGCAAGCATCGCCGGC 708
Db      661 CACGCCGGGCCCCCATCGCCCCCGGGCCAGATGCGCGAGCGCCCGGCAAGCATCGCCGGC 720
QY      709 ACCACGAGCACCCCTGCAAGAGCAGATGCGCTGATGACCAAGCAACCCCGCATCCCGTG 768
Db      721 ACCACGAGCACCCCTGCAAGAGCAGATGCGCTGATGACCAACCAACCCCGCATCCCGTG 780
QY      769 GGGGACATCTACAAGCGGTGATCATCTGGGCTTGAAACAAGATCGTGCGGATGTACAGC 828
Db      781 GGGGAGATCTACAAGCGGTGATCATCTGGGCTTGAAACAAGATCGTGCGGATGTACAGC 840
QY      829 CCCGTGAGCATCTTGACATCAAGCAGGGCCCCCAAGAGCCCTTCCGCACTACGTGAC 888
Db      841 CCCACGAGCATCTTGACATCCGCGCAGGGCCCCCAAGAGCCCTTCCGCACTACGTGAC 900
QY      889 CGCTTCTTCAAGACCTTGCGCGCCGAGAGCACCAGAGGAGTGAAGAACTGATGACC 948
Db      901 CGCTTCTTCAAGACCTTGCGCGCTGAGCAGCCAGCAGACGTGAAGAACTGATGACC 960
QY      949 GACACCTGCTGTGTCAGAACGCCAACCCCGACTGCAAGACCATCTGCGCTCTGGC 1008
Db      961 GAGACCTGCTGTGTCAGAACGCCAACCCCGACTGCAAGACCATCTGAAAGCTCTGGC 1020
QY      1009 CCCGGCGCCAGCCTGAGAGAGATGATGACCGCTGCGAGGGCGTGGCGGCCCAAGCCAC 1068
Db      1021 CCCGGCGCCACCTGAGAGAGATGATGACCGCTGCGAGGGCGTGGCGGCCCAAGCCAC 1080
QY      1069 AAGGCCCGCGTGTGCGCGAGGCGAGTGAAGCCAGGCCAACCA-----CCAGGTGATGATG 1122
Db      1081 AAGGCCCGCGTGTGCGCGAGGCGAGTGAAGCCAGGTGACGAACCCGGCGAACCATGATG 1140
QY      1123 CAGAAGAGCAACTTCAAGGGCCCCCGCGCATGTCTCAAGTGTCTTCAACTGCGGCAAGAG 1182
Db      1141 CAGCGCGGCAACTTCCGCAACCAAGCGGAAAGACCGTCAAGTGTCTTCAACTGCGGCAAGAG 1200
QY      1183 GGGCACAATGCGCCGCAACTGCGCGCGCCCCCGCAAGAAAGGCTGCTGGAAGTGGCGCAAG 1242
Db      1201 GGGCACAACCGCAAGAACTGCGCGCGCCCCCGCAAGAAAGGCTGCTGCGCGCGCGC 1260
QY      1243 GAGGGCCACCAAGATGAAGAGCTGCAACCGAGCGCCAGGCCCACTTCTTGGGCAAGATCTGG 1302
Db      1261 GAGGGCCACCAAGATGAAGAGCTGCAACCGAGCGCCAGGCCCACTTCTTGGGCAAGATCTGG 1320
QY      1303 CCCAGCCACAAGGGCGCGCGCGCAACTTCTGCAAGCGCGCCCGAGCCCAACCGCGCCC 1362
Db      1321 CCCAGCTACAAGGGCGCGCGCGCAACTTCTGCAAGCGCGCCCGAGCCCAACCGCGCCC 1380
QY      1363 CCCGCCGAGAGCTTCCGCTT-----GAGGAGACCAACCCCGCGCGCAAGCAGAG--- 1413
Db      1381 CCCGAGGAGAGCTTCCGCTTCCGCGAGAGAGAAAGCAACCCCAAGCAAGAGAGCCC 1440
QY      1414 ---AGCAAGGACCGCGAGACCCCTGACCAAGCTTGAAGAGCTTGGCAACGACCCCTG 1470
Db      1441 ATCGACAAGGAGCTGTACCCCTGACCAAGCCTGCGGAGCTGTTCGGCAACGACCCCAAGC 1500
QY      1471 AGCCAGTAA 1479
Db      1501 AGCCAGTAA 1509
```

Search completed: March 12, 2005, 21:18:16  
Job time : 6607.56 secs

THIS PAGE BLANK (u9870)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 11:36:50 ; Search time 839.981 Seconds  
(without alignments)  
10423.210 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 1479

Sequence: 1 atggcgccgcgcgcacgcat.....acgacccctgagccagtaa 1479

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: \_geneseqn1980s:\*  
2: \_geneseqn1990s:\*  
3: \_geneseqn2000s:\*  
4: \_geneseqn2001as:\*  
5: \_geneseqn2001bs:\*  
6: \_geneseqn2002as:\*  
7: \_geneseqn2002bs:\*  
8: \_geneseqn2003as:\*  
9: \_geneseqn2003bs:\*  
10: \_geneseqn2003cs:\*  
11: \_geneseqn2003ds:\*  
12: \_geneseqn2004as:\*  
13: \_geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1479	100.0	1479	3	AAAS1609
2	1479	100.0	1479	6	AA144548
3	1479	100.0	1479	6	ABL39953
4	1479	100.0	1479	12	ADM73758
5	1477.4	99.9	3162	10	ADCl3239
6	1477.4	99.9	3462	10	ADCl3237
7	1477.4	99.9	4419	8	ACA03523
8	1477.4	99.9	4419	10	ADCl3240
9	1477.4	99.9	4615	10	ADCl3257
10	1477.4	99.9	4702	10	ADCl3259
11	1477.4	99.9	4716	8	ACA03522
12	1477.4	99.9	4716	10	ADCl3238
13	1475.8	99.8	2742	8	ACA03524
14	1475.8	99.8	2742	10	ADCl3241
15	1475.8	99.8	3930	10	ADCl3230
16	1475.8	99.8	3930	10	ADCl3231
17	1475.8	99.8	3930	10	ADCl3232
18	1475.8	99.8	5145	8	ACA03521
19	1475.8	99.8	5145	10	ADCl3233
20	1474.8	99.7	4713	8	ACA03592

21	1474.8	99.7	4713	10	ADCl3280	AdCl3280 DNA of HI
22	1473.8	99.6	5184	8	ACA03591	Aca03591 Synthetic
23	1473.8	99.6	5184	10	ADCl3279	AdCl3279 DNA of HI
24	1472.4	99.6	2742	8	ACA03590	Aca03590 Synthetic
25	1472.4	99.6	2742	10	ADCl3278	AdCl3278 DNA of HI
26	1463	98.9	1479	3	AAAS1625	Aaas1625 HIV codon
27	1463	98.9	1479	6	ABL39957	Ab139957 Synthetic
28	1463	98.9	1479	12	ADM73762	Adm73762 HIV-1 pol
29	1461.4	98.8	1479	6	AA144552	Aa144552 HIV-1 p55
30	1424.4	96.3	4546	10	ADCl3255	AdCl3255 DNA of HI
31	1347.4	91.1	4423	10	ADCl3256	AdCl3256 DNA of HI
32	1315.2	88.9	3531	10	ADCl3234	AdCl3234 DNA of HI
33	1315.2	88.9	3537	10	ADCl3236	AdCl3236 DNA of HI
34	1315.2	88.9	3538	10	ADCl3235	AdCl3235 DNA of HI
35	1288.8	87.1	1491	6	ABL40020	Ab140020 Synthetic
36	1288.8	87.1	1491	12	ADM73833	Adm73833 HIV-1 pol
37	1288	87.1	1494	6	ABL39972	Ab139972 Synthetic
38	1288	87.1	1494	12	ADM73785	Adm73785 HIV-1 pol
39	1276.8	86.3	1509	3	AAAS1610	Aaas1610 HIV synth
40	1276.8	86.3	1509	6	AA144549	Aa144549 HIV-1 p55
41	1276.8	86.3	1509	6	ABL39954	Ab139954 Synthetic
42	1276.8	86.3	1509	12	ADM73759	Adm73759 HIV-1 pol
43	1262.4	85.4	1509	6	ABL39958	Ab139958 Synthetic
44	1262.4	85.4	1509	12	ADM73763	Adm73763 HIV-1 pol
45	1260.8	85.2	1509	3	AAAS1626	Aaas1626 HIV codon

#### ALIGNMENTS

RESULT 1	
AAAS1609	
ID	AAAS1609 standard; DNA; 1479 BP.
XX	
AC	AAAS1609;
XX	
DT	31-OCT-2000 (first entry)
XX	
DE	HIV synthetic Gag polynucleotide.
XX	
KW	Gag: expression cassette; antigenic; type C; HIV; Env; synthetic;
KW	DNA immunization; packaging cell line; antigen presentation; ss.
XX	
OS	Human immunodeficiency virus; type C strain AF110965.
XX	
FH	Synthetic.
FT	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1479
FT	/product= "Synthetic Gag"
FT	/note= "Codon usage pattern was modified and inhibitory
FT	elements (INS) and RRE sites were inactivated resulting
FT	in improved expression"
XX	
PN	WO200039304-A2.
XX	
PD	06-JUL-2000.
XX	
PF	30-DEC-1999; 99WO-US031273.
XX	
PR	31-DEC-1998; 98US-0114495P.
PR	01-SEP-1999; 99US-0152195P.
XX	
PA	(CHIR ) CHIRON CORP.
XX	
PI	Barnett S, Zur Megede J;
XX	
DR	WPI; 2000-452401/39.
DR	P-PSDB; AAY96943.
XX	
PT	Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env
PT	polypeptide and the polypeptide useful for immunizing a mammal especially
PT	human against HIV.

XX Claim 2; Page 92-93; 113pp; English.  
PS  
XX Expression cassettes comprising a polynucleotide encoding antigenic type  
CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
CC in DNA immunization, generation of packaging cell lines and production of  
CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
CC cassettes exhibit increased potency for induction of cytotoxic T-  
CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-  
CC assemble into non-infectious virus-like particles which are used as a  
CC matrix for the proper presentation of an antigen entrapped or associated  
CC to the immune system of the host  
XX

SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;

Query Match 100.0%; Score 1479; DB 3; Length 1479;  
Best Local Similarity 100.0%; Pred. No. 3.2e-185;  
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCCCGCGCCACATCTCGCGGGCGGCAAGCTGGACGCGCTGGAGCGCATCCGC 60  
DB 1 ATGGGCGCCCGCGCCACATCTCGCGGGCGGCAAGCTGGACGCGCTGGAGCGCATCCGC 60  
QY 61 CTGCGCGCCCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGGCCAGCGCCGAG 120  
DB 61 CTGCGCGCCCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGGCCAGCGCCGAG 120  
QY 121 CTGAGAGAAGTTCGCTTGAACCCCGCTGTGTGAAGCAAGCGAGGCTGCAAGCAGATC 180  
DB 121 CTGAGAGAAGTTCGCTTGAACCCCGCTGTGTGAAGCAAGCGAGGCTGCAAGCAGATC 180  
QY 181 ATCCGCGCAGCTGCAACCCCGCTGTGCAAGCGGCAAGAGTGAAGAGCTGTTCAC 240  
DB 181 ATCCGCGCAGCTGCAACCCCGCTGTGCAAGCGGCAAGAGTGAAGAGCTGTTCAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAAGAGAGTGAAGTCCGCGCAACCAAGAGGCGC 300  
DB 241 ACCGTGGCCACCTGTACTGCGTGCAAGAGAGTGAAGTCCGCGCAACCAAGAGGCGC 300  
QY 301 CTGCAACAAGATCGAGAGAGAGCAACAAGTGCAGAGAGATCCAAGAGGCGCGGCC 360  
DB 301 CTGCAACAAGATCGAGAGAGAGCAACAAGTGCAGAGAGATCCAAGAGGCGCGGCC 360  
QY 361 GCCGCAAGGCGCAAGGTGAAGCCAGAACTAACCCCATGCTGAGAACTTGCAGGCGCAGATG 420  
DB 361 GCCGCAAGGCGCAAGGTGAAGCCAGAACTAACCCCATGCTGAGAACTTGCAGGCGCAGATG 420  
QY 421 GTGCAACAGGCGCATCAACCCCGCGCACTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480  
DB 421 GTGCAACAGGCGCATCAACCCCGCGCACTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480  
QY 481 GCCTTCAGCCCGCGAGGTGATCCCAATGTTCAACGCGCTGAGCGAGGCGCCACCCCGCAG 540  
DB 481 GCCTTCAGCCCGCGAGGTGATCCCAATGTTCAACGCGCTGAGCGAGGCGCCACCCCGCAG 540  
QY 541 GACCTGAACACGATGTTGAACAACCGTGGGCGGCGCAAGCGCGCATGCAAGATGCTGAAG 600  
DB 541 GACCTGAACACGATGTTGAACAACCGTGGGCGGCGCAAGCGCGCATGCAAGATGCTGAAG 600  
QY 601 GACACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCGGTGCAACCGCGCGCGCGCGCC 660  
DB 601 GACACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCGGTGCAACCGCGCGCGCGCGCC 660  
QY 661 ATGCGCGCGCGCGAGATGCGCGAGCGCGCGCGCGCGAGCATCGCGCGCGCAACCAAGCAAC 720  
DB 661 ATGCGCGCGCGCGAGATGCGCGAGCGCGCGCGCGCGAGCATCGCGCGCGCAACCAAGCAAC 720  
QY 721 CTGCAAGAGAGAGATGCGCTGATGACCAACACCCCGCATCCCGGTGGGCGAGCATCTAC 780  
DB 721 CTGCAAGAGAGAGATGCGCTGATGACCAACACCCCGCATCCCGGTGGGCGAGCATCTAC 780  
QY 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC 840  
DB 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC 840

DB 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC 840  
QY 841 CTGACATCAAGAGAGGCGCGCGCAAGAGAGCCCTTCGCGACTACGTTGAGCCGCTTCTCAAG 900  
DB 841 CTGACATCAAGAGAGGCGCGCGCAAGAGAGCCCTTCGCGACTACGTTGAGCCGCTTCTCAAG 900  
QY 901 ACCCTGCGCGCGAGAGAGAGCAACCAAGAGGTGAAGAATGATGACCGACACCTGCTG 960  
DB 901 ACCCTGCGCGCGAGAGAGAGCAACCAAGAGGTGAAGAATGATGACCGACACCTGCTG 960  
QY 961 GTGAGAAACGCAACCCCGACTGCAAGAACCATCTGCGCGCTGCGCGCGCGCGCGCAGC 1020  
DB 961 GTGAGAAACGCAACCCCGACTGCAAGAACCATCTGCGCGCTGCGCGCGCGCGCGCAGC 1020  
QY 1021 CTGAGAGAGATGATGACCGCGCTGCGAGGCGGTGGGCGGCGCGCGCGCGCGCGCTG 1080  
DB 1021 CTGAGAGAGATGATGACCGCGCTGCGAGGCGGTGGGCGGCGCGCGCGCGCGCGCTG 1080  
QY 1081 CTGCGCGAGGCGATGAGCCAGGCGCAACACCAAGCGTATGATGACGAAGAACTTCAAG 1140  
DB 1081 CTGCGCGAGGCGATGAGCCAGGCGCAACACCAAGCGTATGATGACGAAGAACTTCAAG 1140  
QY 1141 GCGCGCGCGCGCGATGCTCAAGTGTCTTCAACTGCGCGCAAGAGAGGCGCACATGCGCGCAAC 1200  
DB 1141 GCGCGCGCGCGCGATGCTCAAGTGTCTTCAACTGCGCGCAAGAGAGGCGCACATGCGCGCAAC 1200  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGAGGCGCAACAGATGAAG 1260  
DB 1201 TGCGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGAGGCGCAACAGATGAAG 1260  
QY 1261 GACTGACCGAGCGCGAGCGCAACTTCTGCGCAAGATCTGGCCAGACCAAGAGGCGCGC 1320  
DB 1261 GACTGACCGAGCGCGAGCGCAACTTCTGCGCAAGATCTGGCCAGACCAAGAGGCGCGC 1320  
QY 1321 CCGCGCAACTTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCCGC 1380  
DB 1321 CCGCGCAACTTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCCGC 1380  
QY 1381 TTGAGAGAGAGCAACCCCGCGCGCAAGAGAGAGCAAGGACCGGAGACCTGACAGC 1440  
DB 1381 TTGAGAGAGAGCAACCCCGCGCGCAAGAGAGAGCAAGGACCGGAGACCTGACAGC 1440  
QY 1441 CTGAAGAGCGCTGTCGCAAGCAACCCCGTGAAGCAAGTAA 1479  
DB 1441 CTGAAGAGCGCTGTCGCAAGCAACCCCGTGAAGCAAGTAA 1479

RESULT 2  
AAL44548  
ID AAL44548 standard; DNA; 1479 BP.  
XX  
AC AAL44548;  
XX  
DT 29-AUG-2003 (revised)  
DT 08-NOV-2002 (first entry)  
XX  
DE HIV-1 p55gag polypeptide coding sequence 1.  
XX  
XX HIV; ds; vaccine; gene; immune response; microparticle;  
KW adsorbent surface; poly(alpha-hydroxy acid); poly(hydroxy butyric acid;  
KW polycaprolactone; polyorthoester; polycyanacrylate; detergent;  
KW submicron emulsion; viral infection; bacterial infection;  
KW parasitic infection; HIV-1 p55gag polypeptide.  
OS Human immunodeficiency virus 1.  
XX  
PN WO200226209-A2.  
XX  
PD 04-APR-2002.  
XX  
PF 28-SEP-2001; 2001WO-US030540.  
XX  
PR 28-SEP-2000; 2000US-0236105P.

PR 30-AUG-2001, 2001US-0315905P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
PI O'hagan D, Otten G, Donnelly JJ, Polo JM, Barnett S, Singh M,  
PI Ulmer J, Dubensky TW;  
XX  
DR WPI; 2002-519084/55.  
XX  
PT A microparticle to which a biologically active macromolecule is adsorbed,  
PT for use as a vaccine composition to treat viral, bacterial or parasitic  
PT infections, comprises a polymer microparticle, a detergent and a  
PT submicron emulsion.  
XX  
XX  
PS Claim 72; Fig 1; 100pp; English.  
XX  
CC The invention relates to a method of raising an immune response in a host  
CC animal. The method of the invention comprises administering a  
CC microparticle that has an adsorbent surface to which a first biologically  
CC active macromolecule (e.g. a nucleic acid) has been adsorbed. The  
CC microparticle comprises a polymer microparticle of poly(alpha-hydroxy  
CC acid), a polyhydroxy butyric acid, a polycaprolactone, a polyorthoester,  
CC a polycyanacrylate, a detergent, and submicron emulsion. The method/  
CC microparticle of the invention is useful for immunising a host animal  
CC against viral, bacterial or parasitic infections. The present DNA  
CC sequence encodes a HIV-1 p55gag polypeptide. (Updated on 29-AUG-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 1479; DB 6; Length 1479;  
Best Local Similarity 100.0%; Pred. No. 3.2e-185;  
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ATGGGGCGCCGCGCGCAGCATCTCGCGCGCGCGCAAGCTGGACGCGCTGGAGCGCATCCGC 60  
Db 1 ATGGGGCGCCGCGCGCAGCATCTCGCGCGCGCGCAAGCTGGACGCGCTGGAGCGCATCCGC 60  
  
QY CTGCGCCCCGCGCGCAAGAAGTGTACATGATGAAGCAACTGTGTGGCCAGCGCGAG 120  
Db CTGCGCCCCGCGCGCAAGAAGTGTGTACATGATGAAGCAACTGTGTGGCCAGCGCGAG 120  
  
QY 121 CTGAGAAGTTGCGCCTGAACCCCGCCTGTGAGACCGGACGAGGCTGAAGAGCCTGTTCAAC 180  
Db 121 CTGAGAAGTTGCGCCTGAACCCCGCCTGTGAGACCGGACGAGGCTGAAGAGCCTGTTCAAC 180  
  
QY 181 ATCCGCCAGCTGCACCCCGCCTGAGACCGGACGAGGAGCTGAAGAGCCTGTTCAAC 240  
Db 181 ATCCGCCAGCTGCACCCCGCCTGAGACCGGACGAGGAGCTGAAGAGCCTGTTCAAC 240  
  
QY 241 ACCGTGGCCACCTGTACTGCGTGACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTACTGCGTGACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300  
  
QY 301 CTGACAAGATCGAGAGGAGGACGACAAGTGCACGACGAGAAGATCGAGAGGCCGCGC 360  
Db 301 CTGACAAGATCGAGAGGAGGAGGACGACAAGTGCACGACGAGAAGATCGAGAGGCCGCGC 360  
  
QY 361 GCCGACAAGGGCAAGGTGAGCCGACAATAACCCCATGTGAGAACTTCAGGGGCCAGATG 420  
Db 361 GCCGACAAGGGCAAGGTGAGCCGACAATAACCCCATGTGAGAACTTCAGGGGCCAGATG 420  
  
QY 421 GTGCAACGAGGCCATCAGCCCCCGCACCTGAACGCGCTGGGTGAAGGTGATCGAGAGAAG 480  
Db 421 GTGCAACGAGGCCATCAGCCCCCGCACCTGAACGCGCTGGGTGAAGGTGATCGAGAGAAG 480  
  
QY 481 GCCTTCAGCCCCCGAGGTGATCCCGCATGTTCAACGCGCCTGAGCGAGGGCGCCACCCCCAG 540  
Db 481 GCCTTCAGCCCCCGAGGTGATCCCGCATGTTCAACGCGCCTGAGCGAGGGCGCCACCCCCAG 540  
  
QY 541 GACCTGAACAAGATGTTGAACAACCGTGGGGCGGCCACGAGCCGCGCATGAGATGCTGAAG 600  
Db 541 GACCTGAACAAGATGTTGAACAACCGTGGGGCGGCCACGAGCCGCGCATGAGATGCTGAAG 600

QY 601 GACACCATCAACGAGAGGCGCGCGAGTGGACCGCGTGCAACCCCGTGACGCGCGGCC 660  
Db 601 GACACCATCAACGAGAGGCGCGCGAGTGGACCGCGTGCAACCCCGTGACGCGCGGCC 660  
  
QY 661 ATGCCCCCGGCGAGATGCGCGAGCGCGCGCGGAGCGACATCGCGCGCACCAAGAGACC 720  
Db 661 ATGCCCCCGGCGAGATGCGCGAGCGCGCGCGGAGCGACATCGCGCGCACCAAGAGACC 720  
  
QY 721 CTGAGAGAGCAGATCGCCTGATGACCAACCCCCCATCCCGTGGGCGACATCTAC 780  
Db 721 CTGAGAGAGCAGATCGCCTGATGACCAACCCCCCATCCCGTGGGCGACATCTAC 780  
  
QY 781 AAGCGGTGATCATCTCTGGGCTGAAACAAGATCGTGGGATGTACAGCCCCGTGACATC 840  
Db 781 AAGCGGTGATCATCTCTGGGCTGAAACAAGATCGTGGGATGTACAGCCCCGTGACATC 840  
  
QY 841 CTGACATCAAGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 900  
Db 841 CTGACATCAAGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 900  
  
QY 901 ACCCTGCGCGCGGACGACGAGACCAAGAGGTGAAGAAGTGAATGATGACCGACACCTGCTG 960  
Db 901 ACCCTGCGCGCGGACGACGAGACCAAGAGGTGAAGAAGTGAATGATGACCGACACCTGCTG 960  
  
QY 961 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGCCCCCGCGCGACG 1020  
Db 961 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGCCCCCGCGCGACG 1020  
  
QY 1021 CTGAGAGAGATGATGACCGCCTGCGCAGGCGCGTGGGCGGCCCGACCAAGGCGCGGTG 1080  
Db 1021 CTGAGAGAGATGATGACCGCCTGCGCAGGCGCGTGGGCGGCCCGACCAAGGCGCGGTG 1080  
  
QY 1081 CTGCGCGAGGCGGATGAGCCAGGCGCAACACAGCGTGATGATGCAAGAGCAACTTCAAG 1140  
Db 1081 CTGCGCGAGGCGGATGAGCCAGGCGCAACACAGCGTGATGATGCAAGAGCAACTTCAAG 1140  
  
QY 1141 GGGCCCCCGGCGCATGCTCAAGTGTCTCAACTGCGGCGAAGAGGCGCACATCGCGCGAAC 1200  
Db 1141 GGGCCCCCGGCGCATGCTCAAGTGTCTCAACTGCGGCGAAGAGGCGCACATCGCGCGAAC 1200  
  
QY 1201 TGCCGCGCCCCCGCAAGAGGCTGTGGAAGTGGCGCAAGAGGCGCACACAGATGAAG 1260  
Db 1201 TGCCGCGCCCCCGCAAGAGGCTGTGGAAGTGGCGCAAGAGGCGCACACAGATGAAG 1260  
  
QY 1261 GACTGCACCGAGGCGCAGGCGCAACTTCTGCGCAAGATCTGGCCAGCCACCAAGGCGCGC 1320  
Db 1261 GACTGCACCGAGGCGCAGGCGCAACTTCTGCGCAAGATCTGGCCAGCCACCAAGGCGCGC 1320  
  
QY 1321 CCGGCAACTTCTGTCAGAGCGCGCCGAGCCACCGCCCCCGCGGAGAGCTTCGCG 1380  
Db 1321 CCGGCAACTTCTGTCAGAGCGCGCCGAGCCACCGCCCCCGCGGAGAGCTTCGCG 1380  
  
QY 1381 TTCAGAGAGACCAACCCCGCGCAAGAGCAGAGCAAGAGCAGACCTGAACAGC 1440  
Db 1381 TTCAGAGAGACCAACCCCGCGCAAGAGCAGAGCAAGAGCAGACCTGAACAGC 1440  
  
QY 1441 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAGTAA 1479  
Db 1441 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAGTAA 1479  
  
RESULT 3  
ABL39953  
ID ABL39953 standard; DNA; 1479 BP.  
XX  
AC ABL39953;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Synthetic gag polynucleotide sequence SEQ ID NO:3.  
XX  
KW Human immunodeficiency virus type C; antigenic HIV type C protein;

KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
KM immunostimulant; gene therapy; gene; ds.  
XX  
OS Human immunodeficiency virus; type C.  
OS Synthetic.  
XX WO200204493-A2.  
PN  
PD 17-JAN-2002.  
XX  
XX 05-JUL-2001; 2001WO-US021241.  
PF  
XX 05-JUL-2000; 2000US-00610313.  
PR  
XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
PI  
XX MPI; 2002-154920/20.  
DR  
XX  
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunization or generation of packaging cell  
PT lines, particularly in gene therapy.  
XX  
XX Example 1; Fig 1; 233pp; English.  
PS  
XX The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (I). (I) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC Type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
XX Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 1479; DB 6; Length 1479;  
Best Local Similarity 100.0%; Pred. No. 3.2e-185;  
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGGCGCCCGCCAGCATCTCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
DB 1 ATGGGGCGCCCGCCAGCATCTCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60

QY 61 CTGCGCGCCCGCGCAAGAAGTGTACATGATGAAGCACTGTGTGTGGCGCGCGAG 120  
DB 61 CTGCGCGCCCGCGCAAGAAGTGTACATGATGAAGCACTGTGTGTGGCGCGCGAG 120

QY 121 CTGGAAGAAGTGTGCTGAAACCCCGGCTGTGAGAGACCAAGGAGGCTGCAAGCATC 180  
DB 121 CTGGAAGAAGTGTGCTGAAACCCCGGCTGTGAGAGACCAAGGAGGCTGCAAGCATC 180

QY 181 ATCCGCCAGCTGCAACCCCGCTGCAAGACCGGCAAGGAGCTGAAGAGCTTCAAC 240  
DB 181 ATCCGCCAGCTGCAACCCCGCTGCAAGACCGGCAAGGAGCTGAAGAGCTTCAAC 240

QY 241 ACCGTGGCCACCTGTACTGCTGTGACAGAGATCGAGGTCCGCGACACCAAGGAGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGCTGTGACAGAGATCGAGGTCCGCGACACCAAGGAGGCC 300

QY 301 CTGGAACAAGATCGAGAGAGCAACAAGTGTCCAGAGATCCAGCAGGCGCGAGGCC 360  
DB 301 CTGGAACAAGATCGAGAGAGCAACAAGTGTCCAGAGATCCAGCAGGCGCGAGGCC 360

QY 361 GCCGACAAGGGCAAGGTGAGCGCAAGTACCCCATCTGTGCAAGAACTGCAAGGCCAGATG 420  
DB 361 GCCGACAAGGGCAAGGTGAGCGCAAGTACCCCATCTGTGCAAGAACTGCAAGGCCAGATG 420

QY 421 GTGACCAAGGCCATCAGCCCCCGCACCTGTAAAGCTGTGTAAGGTATCGAGAGAAG 480  
DB 421 GTGACCAAGGCCATCAGCCCCCGCACCTGTAAAGCTGTGTAAGGTATCGAGAGAAG 480

QY 481 GCGTTAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCGCAACCCCCAG 540  
DB 481 GCGTTAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCGCAACCCCCAG 540

QY 541 GACCTGAACAGATGTTGAACAACGTGGCGGCGCACAGCGCCCATGACAGATGCTGAAG 600  
DB 541 GACCTGAACAGATGTTGAACAACGTGGCGGCGCACAGCGCCCATGACAGATGCTGAAG 600

QY 601 GACACCATCAACGAGAGCGCGCGAGTGGGACCGCGTGCAACCCCGTGACCGCGGCC 660  
DB 601 GACACCATCAACGAGAGCGCGCGAGTGGGACCGCGTGCAACCCCGTGACCGCGGCC 660

QY 661 ATCGCCCCCGGCGAGATGCGCGAGCCCCCGCGGACGACATCGCCGACCAACGACACC 720  
DB 661 ATCGCCCCCGGCGAGATGCGCGAGCCCCCGCGGACGACATCGCCGACCAACGACACC 720

QY 721 CTGAGAGAGATCGCTGTGATGACCAAGCAACCCCGCATCCCGTGGCGCATCTAC 780  
DB 721 CTGAGAGAGATCGCTGTGATGACCAAGCAACCCCGCATCCCGTGGCGCATCTAC 780

QY 781 AAGCGGTGATCATCTTGGGCTGAACAAGATCGTGCAGATGTACAGCCCCGTGAGCATC 840  
DB 781 AAGCGGTGATCATCTTGGGCTGAACAAGATCGTGCAGATGTACAGCCCCGTGAGCATC 840

QY 841 CTGACATCAAGCAGAGCGCCCAAGAGAGCCCTTCGCGATACGTGACCGCTTCTTCAAG 900  
DB 841 CTGACATCAAGCAGAGCGCCCAAGAGAGCCCTTCGCGATACGTGACCGCTTCTTCAAG 900

QY 901 ACCCTGCGCGCGAGCAGAGCAACCCAGAGGTGAAGAACTGATGACCGACACCTGCTG 960  
DB 901 ACCCTGCGCGCGAGCAGAGCAACCCAGAGGTGAAGAACTGATGACCGACACCTGCTG 960

QY 961 GTGCAGAACGCCAACCCCGACTGCAAGAACATCTGCGGCTCTGCGCCCGCGCGCAGC 1020  
DB 961 GTGCAGAACGCCAACCCCGACTGCAAGAACATCTGCGGCTCTGCGCCCGCGCGCAGC 1020

QY 1021 CTGAGAGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCCAGCAAGGCCGCGTG 1080  
DB 1021 CTGAGAGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCCAGCAAGGCCGCGTG 1080

QY 1081 CTGGCGGAGCGATGAGCGCGCAACCAACAGCGTATGATGCAAGAGCAACTTCAAG 1140  
DB 1081 CTGGCGGAGCGATGAGCGCGCAACCAACAGCGTATGATGCAAGAGCAACTTCAAG 1140

QY 1141 GGGCCCCGGCGCATGTCAAGTGTCTCAACTGCGGCAAGAGGGGCCAATCGCCCGCAAC 1200  
DB 1141 GGGCCCCGGCGCATGTCAAGTGTCTCAACTGCGGCAAGAGGGGCCAATCGCCCGCAAC 1200

QY 1201 TGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCGGCAAGGAGGCGCAACGATGAAG 1260  
DB 1201 TGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCGGCAAGGAGGCGCAACGATGAAG 1260

QY 1261 GACTGACCGAGCGCGCAAGCACTTCTGGGGCAAGATCTGGCCCAAGAGGCGCGC 1320  
DB 1261 GACTGACCGAGCGCGCAAGCACTTCTGGGGCAAGATCTGGCCCAAGAGGCGCGC 1320

QY 1321 CCGGCAACTTCTGCAAGCGCGCGCGACCGCCCGCCCCCGCGAGAGCTTCCGC 1380  
DB 1321 CCGGCAACTTCTGCAAGCGCGCGCGACCGCCCGCCCCCGCGAGAGCTTCCGC 1380

QY 1381 TTGAGAGAACCAACCCCGCGCAAGAGCAAGAGAGCAAGAACCGGAGACCTGACCAAGC 1440  
DB 1381 TTGAGAGAACCAACCCCGCGCAAGAGCAAGAGAGCAAGAACCGGAGACCTGACCAAGC 1440

QY 1441 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAGTAA 1479  
DB 1441 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAGTAA 1479

```
RESULT 4
ADM73758
ID ADM73758 standard; DNA; 1479 BP.
XX
AC ADM73758;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #1.
XX
KM HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGEL/) ENGELBRECHT S.
PA (RENS/) RENSBURG B J V.
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Example 1; SEQ ID NO 3; 160bp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;
Query Match 100.0%; Score 1479; DB 12; Length 1479;
Best Local Similarity 100.0%; Pred. No. 3.2e-185;
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGGCGCCCGCGCCAGCATCTGCGCGCGCGCAAGCTGACGCGCTGGAGCGCATCCGC 60
DB 1 ATGGGCGCCCGCGCCAGCATCTGCGCGCGCGCAAGCTGACGCGCTGGAGCGCATCCGC 60
QY 61 CTGCGCCCCCGCGCGCAAGAGTGCTACATGATGAAGCACTGTGTGGGCCAGCCGCGAG 120
DB 61 CTGCGCCCCCGCGCGCAAGAGTGCTACATGATGAAGCACTGTGTGGGCCAGCCGCGAG 120
QY 121 CTGAGAAAGTTGCGCCCTGAACCCCGGCTGTGAGACCAAGGAGCTGCAAGCAGATC 180
DB 121 CTGAGAAAGTTGCGCCCTGAACCCCGGCTGTGAGACCAAGGAGCTGCAAGCAGATC 180
QY 181 ATCCGCGAGCTGACCCCGCCCTGCAAGCCGCGAGCGAGGAGCTGAAGCCTGTTCAAC 240
DB 181 ATCCGCGAGCTGACCCCGCCCTGCAAGCCGCGAGCGAGGAGCTGAAGCCTGTTCAAC 240
```

```
DB 181 ATCCGCGAGCTGACCCCGCCCTGCAAGCCGCGAGCGAGGAGCTGAAGCCTGTTCAAC 240
QY 241 ACCGTGGCCACCCCTGTACTGCGTGCACGAGAAGATGAGGTCGCGGACACCAAGAGGCC 300
DB 241 ACCGTGGCCACCCCTGTACTGCGTGCACGAGAAGATGAGGTCGCGGACACCAAGAGGCC 300
QY 301 CTGGAACAAGATCGAGGAGGAGCAGAACAAAGTGCCACAGAAAGATTCAGAGGCCGAGGCC 360
DB 301 CTGGAACAAGATCGAGGAGGAGCAGAACAAAGTGCCACAGAAAGATTCAGAGGCCGAGGCC 360
QY 361 GCCGACAAAGGCGAAGGTGAGCCAGAACTACCCCATCTGTGCAAACTTGAGGGCCAGATG 420
DB 361 GCCGACAAAGGCGAAGGTGAGCCAGAACTACCCCATCTGTGCAAACTTGAGGGCCAGATG 420
QY 421 GTGCAACGAGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480
DB 421 GTGCAACGAGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCTGAGCGAGGGCGCCACCCCCAG 540
DB 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCTGAGCGAGGGCGCCACCCCCAG 540
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGGCCACAGGCCGCGCATGCAAGATGCTGAAG 600
DB 541 GACCTGAACACGATGTTGAACACCGTGGCGGCCACAGGCCGCGCATGCAAGATGCTGAAG 600
QY 601 GACACCATCAACGAGAGAGGCCCGCGAGTGGACCGCGTGCAACCCGCTGCAAGCCGCCCC 660
DB 601 GACACCATCAACGAGAGAGGCCCGCGAGTGGACCGCGTGCAACCCGCTGCAAGCCGCCCC 660
QY 661 ATGCCCCCGGCGCAGATGCGCGAGCCCGCGCGCGAGCATGCGCGGACCAACAGACACC 720
DB 661 ATGCCCCCGGCGCAGATGCGCGAGCCCGCGCGCGAGCATGCGCGGACCAACAGACACC 720
QY 721 CTGCAAGAGCAGATGCGCTGATGACCAAGCAACCCCATCCCGTGGGCGCACTCTAC 780
DB 721 CTGCAAGAGCAGATGCGCTGATGACCAAGCAACCCCATCCCGTGGGCGCACTCTAC 780
QY 781 AAGCGGTGATCATCTGGGCTGGAACAAGATCGTGGATGTACAGCCCGTGAGCATC 840
DB 781 AAGCGGTGATCATCTGGGCTGGAACAAGATCGTGGATGTACAGCCCGTGAGCATC 840
QY 841 CTGACATCAAGCGAGGCCCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900
DB 841 CTGACATCAAGCGAGGCCCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900
QY 901 ACCCTGCGCGCGAGCAGAGACCAAGAGGTGAAGAACTGATGATCCGACACCTGCTG 960
DB 901 ACCCTGCGCGCGAGCAGAGACCAAGAGGTGAAGAACTGATGATCCGACACCTGCTG 960
QY 961 GTGCAAGACGCAACCCCGACTGCAAGACCATCTGCGCGCTTCTGGCCCCGCGCAGC 1020
DB 961 GTGCAAGACGCAACCCCGACTGCAAGACCATCTGCGCGCTTCTGGCCCCGCGCAGC 1020
QY 1021 CTGAGAGAGATGATACCGCCTGCGCAGGGCGTGGGGGGCCCCCAGCCACAAGGCCGCGTG 1080
DB 1021 CTGAGAGAGATGATACCGCCTGCGCAGGGCGTGGGGGGCCCCCAGCCACAAGGCCGCGTG 1080
QY 1081 CTGCGCGAGGCGCATGAGCCAGGCCAACACCAAGCGTGATGATGCAAGAAGCACTTCAAG 1140
DB 1081 CTGCGCGAGGCGCATGAGCCAGGCCAACACCAAGCGTGATGATGCAAGAAGCACTTCAAG 1140
QY 1141 GGGCCCCGGCGCATCTGTCAAGTCTTCACTGCGGCAAGAGGGCCACATGCGCCGCAAC 1200
DB 1141 GGGCCCCGGCGCATCTGTCAAGTCTTCACTGCGGCAAGAGGGCCACATGCGCCGCAAC 1200
QY 1201 TGCCGCGCCCCCGCAAGAAGGGTGCTGGAATGCGGCAAGAGGGCCACCAAGTGAAG 1260
DB 1201 TGCCGCGCCCCCGCAAGAAGGGTGCTGGAATGCGGCAAGAGGGCCACCAAGTGAAG 1260
QY 1261 GACTGCAACGAGCGCCAGGCCAATTCTGTGGCAAGATCTGGCCAGCCACAAGGGCGCC 1320
DB 1261 GACTGCAACGAGCGCCAGGCCAATTCTGTGGCAAGATCTGGCCAGCCACAAGGGCGCC 1320
```



QY 1141 GGGCCCCGGCATGTCAGTGTCTCAACTGCGGCAAGAGGGCCATCGCCCGCAAC 1200  
DB 1147 GGGCCCCGGCATGTCAGTGTCTCAACTGCGGCAAGAGGGCCATCGCCCGCAAC 1206  
QY 1201 TGCCGGCCCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGAGGGCCACAGATGAAG 1260  
DB 1207 TGCCGGCCCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGAGGGCCACAGATGAAG 1266  
QY 1261 GACTGCACCGAGCGCGCAAGCTTCTGGGCAAGATCTGGCCAGCCACCAAGGGCCGC 1320  
DB 1267 GACTGCACCGAGCGCGCAAGCTTCTGGGCAAGATCTGGCCAGCCACCAAGGGCCGC 1326  
QY 1321 CCCGGCAACTTCTGTGAGAGCGCGCCGAGCCCGCCCCCGCGAGAGCTTCCGC 1380  
DB 1327 CCCGGCAACTTCTGTGAGAGCGCGCCGAGCCCGCCCCCGCGAGAGCTTCCGC 1386  
QY 1381 TTGAGAGAGACCAACCCCGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1387 TTGAGAGAGAGACCAACCCCGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
QY 1441 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAGTAA 1479  
DB 1447 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAGAA 1485

RESULT 6  
ADCI3237  
ID ADCI3237 standard; DNA; 3462 BP.  
XX  
AC ADCI3237;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA of HIV construct GagProtiNarMut\_C SEQ ID NO 16.  
XX  
KW expression cassette; HIV Gag; Env; Int; Nef; p15NaseH; Pol; Tat; Prot;  
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX  
OS Human immunodeficiency virus.  
XX  
PN WO2003004620-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 05-JUL-2002; 2002WO-US021420.  
XX  
PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX  
DR WPI; 2003-221593/21.  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15NaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX  
PS Disclosure; Fig 13; 301bp; English.  
XX  
CC The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15NaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramuscularly, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful

CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX  
SQ Sequence 3462 BP; 790 A; 1198 C; 1068 G; 406 T; 0 U; 0 Other;  
Query Match 99.9%; Score 1477.4; DB 10; Length 3462;  
Best Local Similarity 99.9%; Pred. No. 4.5e-185;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGGGCGCCCGCGCAGCATCTCTGCGCGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
DB 7 ATGGGCGCCCGCGCAGCATCTCTGCGCGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 66  
QY 61 TTGCGCCCCCGGCGCAAGAGTGTCTATGATGAAGCACCTGTGTGGCCAGCGCGAG 120  
DB 67 TTGCGCCCCCGGCGCAAGAGTGTCTATGATGAAGCACCTGTGTGGCCAGCGCGAG 126  
QY 121 CTGAGAGAGTTGCGCTGTAACCCCGGCTGTGAGAGACGAGCGGCTGCAAGCAGATC 180  
DB 127 CTGAGAGAGTTGCGCTGTAACCCCGGCTGTGAGAGACGAGCGGCTGCAAGCAGATC 186  
QY 181 ATCCGCCAGCTGACCCCGGCTGTGAGAGACCGGAGCGAGAGCTGAAGAGCTGTTCAC 240  
DB 187 ATCCGCCAGCTGACCCCGGCTGTGAGAGACCGGAGCGAGAGCTGAAGAGCTGTTCAC 246  
QY 241 ACCGTGGCCACCTGTACTGTGCTGTGACGAGAGATCGAGGTCCGCAACCAAGAGGCC 300  
DB 247 ACCGTGGCCACCTGTACTGTGCTGTGACGAGAGATCGAGGTCCGCAACCAAGAGGCC 306  
QY 301 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGGACAAGGGAAGGTGAGAGCAAACTACCCCATCTGTGAGAGAGAGAGAGAGAGAG 420  
DB 367 GCGGACAAGGGAAGGTGAGAGCAAACTACCCCATCTGTGAGAGAGAGAGAGAGAGAG 426  
QY 421 GTGCAACGAGGCTCAGCCCCCGCAACCTGAAAGCTGTGAGAGAGAGAGAGAGAGAG 480  
DB 427 GTGCAACGAGGCTCAGCCCCCGCAACCTGAAAGCTGTGAGAGAGAGAGAGAGAGAG 486  
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCAACCCCCAG 540  
DB 487 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCAACCCCCAG 546  
QY 541 GACCTGAACAGATGTTGAACACCGTGGCGGCAACGAGCGCCATGAGATGCTGAAG 600  
DB 547 GACCTGAACAGATGTTGAACACCGTGGCGGCAACGAGCGCCATGAGATGCTGAAG 606  
QY 601 GACACCATCAACGAGAGAGGCGCGGAGTGGGACCGCGTGCAACCCCGTGCAACCGCGCCC 660  
DB 607 GACACCATCAACGAGAGAGGCGCGGAGTGGGACCGCGTGCAACCCCGTGCAACCGCGCCC 666  
QY 661 ATCGCCCCCGGCGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
DB 667 ATCGCCCCCGGCGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
QY 721 CTGCAAGAGAGATGCGCTGTGATGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
DB 727 CTGCAAGAGAGATGCGCTGTGATGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786  
QY 781 AAGCGGTGATCATCTCTGGGCTGAACAAGATCTGCGGATGTAAGAGAGAGAGAGAGATC 840  
DB 787 AAGCGGTGATCATCTCTGGGCTGAACAAGATCTGCGGATGTAAGAGAGAGAGAGATC 846  
QY 841 CTGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 847 CTGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906  
QY 901 ACCCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 907 ACCCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966

QY 961 GTGAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGCCCCGGCCAGC 1020  
 DB 967 GTGAGAACGCCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGCCCCGGCCAGC 1026  
 QY 1021 CTGGAGAGATGATGACCGCTCTGCAAGGGCGTGGGGCGCCCGCCAGCCAGCGCGGTG 1080  
 DB 1027 CTGGAGAGATGATGACCGCTCTGCAAGGGCGTGGGGCGCCCGCCAGCCAGCGCGGTG 1086  
 QY 1081 CTGGCCGAGGCGATGAGCCAGGCGCAACACCGCGTGATGATGCAAGAGCACTTCAAG 1140  
 DB 1087 CTGGCCGAGGCGATGAGCCAGGCGCAACACCGCGTGATGATGCAAGAGCACTTCAAG 1146  
 QY 1141 GGGCCCCGGCGCATGTCAGAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGCAAC 1200  
 DB 1147 GGGCCCCGGCGCATGTCAGAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGCAAC 1206  
 QY 1201 TGCGCGCGCCCCCGCGCAAGAGGGCTGCTGGAAGTGGCGGCAAGAGGGCCACCATGAAG 1260  
 DB 1207 TGCGCGCGCCCCCGCGCAAGAGGGCTGCTGGAAGTGGCGGCAAGAGGGCCACCATGAAG 1266  
 QY 1261 GACTGCAACCGAGCGCGCAAGGCGCAACTTCTTGCGCAAGATCTGCGCCAGCCACAGGGCGCG 1320  
 DB 1267 GACTGCAACCGAGCGCGCAAGGCGCAACTTCTTGCGCAAGATCTGCGCCAGCCACAGGGCGCG 1326  
 QY 1321 CCGCGCAACTTCTTGCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
 DB 1327 CCGCGCAACTTCTTGCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386  
 QY 1381 TTGAGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440  
 DB 1387 TTGAGAGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1446  
 QY 1441 CTGAAGAGCTGTGTTGGCAACGACCGCCCTGAGCCAGTAA 1479  
 DB 1447 CTGAAGAGCTGTGTTGGCAACGACCGCCCTGAGCCAGTAA 1485

RESULT 7  
 ACA03523  
 ID ACA03523 standard; DNA; 4419 BP.

AC ACA03523;  
 XX 22-MAY-2003 (first entry)  
 DE Synthetic DNA encoding immunogenic HIV peptide #6.  
 XX Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 OS Synthetic.  
 XX WO2003004657-A1.  
 PN 16-JAN-2003.  
 PD 05-JUL-2002; 2002WO-US021421.  
 PF 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 PA (CHIR ) CHIRON CORP.  
 PI Zur Megele J, Barnett SW, Lian Y;  
 DR WPI; 2003-221602/21.  
 XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C

PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 PS Example 1; Fig 11; 262pp; English.  
 XX The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide

Query Match 99.9%; Score 1477.4; DB 8; Length 4419;  
 Best Local Similarity 99.9%; Pred. No. 4.3e-185;  
 Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
 DB 7 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66  
 QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
 DB 67 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 126  
 QY 121 CTGAGAGAGTTCGCGCTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
 DB 127 CTGAGAGAGTTCGCGCTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186  
 QY 181 ATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
 DB 187 ATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246  
 QY 241 ACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
 DB 247 ACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
 QY 301 CTGAGACAGATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 DB 307 CTGAGACAGATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
 QY 361 GCGGACAAAGGCGAGGTGAGCGCAAGTACCCCATCTGTCAGAACTGTCAGAGGCGCAGATG 420  
 DB 367 GCGGACAAAGGCGAGGTGAGCGCAAGTACCCCATCTGTCAGAACTGTCAGAGGCGCAGATG 426  
 QY 421 GTGACACAGGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 DB 427 GTGACACAGGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486  
 QY 481 GCCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
 DB 487 GCCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546  
 QY 541 GACCTGAACAGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
 DB 547 GACCTGAACAGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606  
 QY 601 GACACATCAACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
 DB 607 GACACATCAACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666  
 QY 661 ATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
 DB 667 ATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726

QY	721	CTGCAGGAGCAGATCGCCTTGATGACCAACCCCCCATCCCGTGGGCGCATCTAC	780
Db	727	CTGCAGGAGCAGATCGCCTTGATGACCAACCCCCCATCCCGTGGGCGCATCTAC	786
QY	781	AAGCGGTGGATCATCTTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCCGCGAGCATC	840
Db	787	AAGCGGTGGATCATCTTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCCGCGAGCATC	846
QY	841	CTGCACATCAAGCAGGGCCCCCAAGAGGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAG	900
Db	847	CTGCACATCAAGCAGGGCCCCCAAGAGGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAG	906
QY	901	ACCCTGCGCGCCGAGCAGAGCACCAGAGGATGAAGAATGTGATGACCGACACCTGCTG	960
Db	907	ACCCTGCGCGCCGAGCAGAGCACCAGAGGATGAAGAATGTGATGACCGACACCTGCTG	966
QY	961	GTGCAGAACGCCCAACCCCGACTGCGAAGACCATCTCGCGGCTCTCGCCCCGCGCCAGC	1020
Db	967	GTGCAGAACGCCCAACCCCGACTGCGAAGACCATCTCGCGGCTCTCGCCCCGCGCCAGC	1026
QY	1021	CTGCAGGAGATGATGACCGCCTTGCCAGAGGCGGTGGGCGGCCCAACCAAGGCGCGGTG	1080
Db	1027	CTGCAGGAGATGATGACCGCCTTGCCAGAGGCGGTGGGCGGCCCAACCAAGGCGCGGTG	1086
QY	1081	CTGGCCGAGGCGGATGAGCCAGGCGCAACACCAGCGTGATGATGACGAAGAGCAATTCAAG	1140
Db	1087	CTGGCCGAGGCGGATGAGCCAGGCGCAACACCAGCGTGATGATGACGAAGAGCAATTCAAG	1146
QY	1141	GCGCCCCCGGCGCATCGTCAAGTGTCTTCAACTGCGGCAAGAGGCGCACATCGCGCGCAAC	1200
Db	1147	GCGCCCCCGGCGCATCGTCAAGTGTCTTCAACTGCGGCAAGAGGCGCACATCGCGCGCAAC	1206
QY	1201	TGCGCGCGCCCCCGGCAAGAGGCGTGTGGAAGTGCGGCAAGAGGCGCACAGATGAAG	1260
Db	1207	TGCGCGCGCCCCCGGCAAGAGGCGTGTGGAAGTGCGGCAAGAGGCGCACAGATGAAG	1266
QY	1261	GACTGCACCCGAGCGGCCAGGCCAACTTCTTGCGCAAGATCTGGCCAGCCACAAGGGCGCG	1320
Db	1267	GACTGCACCCGAGCGGCCAGGCCAACTTCTTGCGCAAGATCTGGCCAGCCACAAGGGCGCG	1326
QY	1321	CCCGGCAACTTCTTGCAAGAGCCCGCGCCGAGCCACCGCCCCCGCGGAGAGCTTCCGC	1380
Db	1327	CCCGGCAACTTCTTGCAAGAGCCCGCGCCGAGCCACCGCCCCCGCGGAGAGCTTCCGC	1386
QY	1381	TTCGAGGAGACCAACCCCCGCGCCAGAAAGCAGAGGCAAGGACCGCGAGACCTTGACCAAGC	1440
Db	1387	TTCGAGGAGACCAACCCCCGCGCCAGAAAGCAGAGGCAAGGACCGCGAGACCTTGACCAAGC	1446
QY	1441	CTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAGTAA 1479	
Db	1447	CTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAGAAA 1485	
RESULT 8			
ID	ADC13240	standard; DNA; 4419 BP.	
AC	ADC13240;		
XX			
DT	18-DEC-2003	(first entry)	
DE	DNA of HIV construct GagRTmutTatRevNef_C SEQ ID NO 19.		
XX			
KW	expression cassette; HIV Gag; Env; Int; Nef; p15RnaeH; Pol; Tat; ProC;		
XX	Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.		
OS	Human immunodeficiency virus.		
XX			
PN	WO2003004620-A2.		
XX			
PD	16-JAN-2003.		
XX			
PF	05-JUL-2002; 2002WO-US021420.		

XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
PI  
XX WPI; 2003-221593/21.  
DR  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX  
XX  
PS Disclosure; Fig 16; 301pp; English.  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramuscosally, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
XX Type C related sequence of the invention.  
XX  
XX Sequence 4419 BP; 980 A; 1582 C; 1371 G; 486 T; 0 U; 0 Other;

Query Match	99.9%;	Score 1477.4;	DB 10;	Length 4419;
Best Local Similarity	99.9%;	Pred. No. 4.3e-185;		
Matches 1478;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps	0			
QY	1	ATGGGCGCCCGCCGACGATCTCGCGCGCGGCAAGCTGACGCTTGGGAGCGCATCCGC	60	
Db	7	ATGGGCGCCCGCCGACGATCTCGCGCGCGGCAAGCTGACGCTTGGGAGCGCATCCGC	66	
QY	61	CTGCGCCCCGGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGGCCAGCCGCGAG	120	
Db	67	CTGCGCCCCGGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGGCCAGCCGCGAG	126	
QY	121	CTGAGAGAGTTCCCTCTGAACCCCGGCTGTCTGAGACCAGCGGCTGCAGCAGATC	180	
Db	127	CTGAGAGAGTTCCCTCTGAACCCCGGCTGTCTGAGACCAGCGGCTGCAGCAGATC	186	
QY	181	ATCCGCCAGCTGCACCCCGCCTGTGACGACCGGCGAGGAGCTGAAGCCTGTTCAAC	240	
Db	187	ATCCGCCAGCTGCACCCCGCCTGTGACGACCGGCGAGGAGCTGAAGCCTGTTCAAC	246	
QY	241	ACCGTGGCCACCCTGTACTGCGTGACGAGAAAGATCGAGTCCGCGACACCAAGAGGCC	300	
Db	247	ACCGTGGCCACCCTGTACTGCGTGACGAGAAAGATCGAGTCCGCGACACCAAGAGGCC	306	
QY	301	CTGGAACAAGATCGAGGAGGAGCAGAACAGTGCCAGCAGAAGATCCAGGCGGAGCC	360	
Db	307	CTGGAACAAGATCGAGGAGGAGCAGAACAGTGCCAGCAGAAGATCCAGGCGGAGCC	366	
QY	361	GCCGACAAAGGGCAAAGTGAGCCAGAACTACCCCATCTGTGAGAACTTGCAAGGGCCAGATG	420	
Db	367	GCCGACAAAGGGCAAAGTGAGCCAGAACTACCCCATCTGTGAGAACTTGCAAGGGCCAGATG	426	
QY	421	GTCGACCAAGGCCATCAGCCCCCGGCAACCCTGAACGCTTGGGTGAAGGTGATCGAGGAGAG	480	
Db	427	GTCGACCAAGGCCATCAGCCCCCGGCAACCCTGAACGCTTGGGTGAAGGTGATCGAGGAGAG	486	
QY	481	GCCTTCAGCCCCCGGAGGTGATCCCATGTTCAACCGCCTGAGCGAGGGCGCCACCCCCAG	540	
Db	487	GCCTTCAGCCCCCGGAGGTGATCCCATGTTCAACCGCCTGAGCGAGGGCGCCACCCCCAG	546	

```
QY 541 GACCTGAACACGATGTTGAACACCCGTTGGGGCGGCCACCAAGGCCCATGCAATGCTGAAG 600
    |||
Db 547 GACCTGAACACGATGTTGAACACCCGTTGGGGCGGCCACCAAGGCCCATGCAATGCTGAAG 606
QY 601 GACACCATCAACGAGAGAGGGCCGCGAGTGGGACCCGCGTGCACCCCGTGCAGCGCGGCC 660
    |||
Db 607 GACACCATCAACGAGAGAGGGCCGCGAGTGGGACCCGCGTGCACCCCGTGCAGCGCGGCC 666
QY 661 ATCGCCCCCGCCAGATGCGCGAGAGCCCCCGCGGAGAGCAATCGCCGCGACCAACCAAGACC 720
    |||
Db 667 ATCGCCCCCGCCAGATGCGCGAGAGCCCCCGCGGAGAGCAATCGCCGCGACCAACCAAGACC 726
QY 721 CTGCAAGAGAGAGATGCGCTTGATGACCAACACCCCCCATCCCGGTGGGCGACATCTAC 780
    |||
Db 727 CTGCAAGAGAGAGATGCGCTTGATGACCAACACCCCCCATCCCGGTGGGCGACATCTAC 786
QY 781 AAGCGGTGATCATCTGGGCTGTAA CAAGATCGTGGGATGTACAGCCCCCGTGAAGCATC 840
    |||
Db 787 AAGCGGTGATCATCTGGGCTGTAA CAAGATCGTGGGATGTACAGCCCCCGTGAAGCATC 846
QY 841 CTGACATCAAGACAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 900
    |||
Db 847 CTGACATCAAGACAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 906
QY 901 ACCCTGCGCGCCGAGACAGACACCAAGAGGTGAAGAACTGATGACCGACACCTGTG 960
    |||
Db 907 ACCCTGCGCGCCGAGACAGACACCAAGAGGTGAAGAACTGATGACCGACACCTGTG 966
QY 961 GTGCAGAAACGCAACCCCGACTGTGCAAGACCATCTGCGCGCTCTCGGCCCGCGCGCAGC 1020
    |||
Db 967 GTGCAGAAACGCAACCCCGACTGTGCAAGACCATCTGCGCGCTCTCGGCCCGCGCGCAGC 1026
QY 1021 CTGAGAGAGATGATGACCGCTGTGCAAGGGCGTGGGGGGCCCCAGACCAAGCGCGCGTG 1080
    |||
Db 1027 CTGAGAGAGATGATGACCGCTGTGCAAGGGCGTGGGGGGCCCCAGACCAAGCGCGCGTG 1086
QY 1081 CTGCGCGAGGCGATGAGCCAGGCCAACCAACGAGCTGATGATGACAGAGAGCAACTTCAAG 1140
    |||
Db 1087 CTGCGCGAGGCGATGAGCCAGGCCAACCAACGAGCTGATGATGACAGAGAGCAACTTCAAG 1146
QY 1141 GGGCCCCGCGCATCTCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATGCGCGCAAC 1200
    |||
Db 1147 GGGCCCCGCGCATCTCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATGCGCGCAAC 1206
QY 1201 TGCGCGCGCCCCCGCAAGAAAGGCTGCTGGAAGTGGCGCAAGGAGGCCACCAAGTGAAG 1260
    |||
Db 1207 TGCGCGCGCCCCCGCAAGAAAGGCTGCTGGAAGTGGCGCAAGGAGGCCACCAAGTGAAG 1266
QY 1261 GACTGCAACCGAGCGCGCAAGCCCAACTTCTTGGGCAAGATCTGCGCACCAAGGGCGCG 1320
    |||
Db 1267 GACTGCAACCGAGCGCGCAAGCCCAACTTCTTGGGCAAGATCTGCGCACCAAGGGCGCG 1326
QY 1321 CCGGCAACTTCTTGCAGAGCGCGCGCGAGGCCACCGCCCCCGCGCGAGAGCTTCCGC 1380
    |||
Db 1327 CCGGCAACTTCTTGCAGAGCGCGCGCGAGGCCACCGCCCCCGCGCGAGAGCTTCCGC 1386
QY 1381 TTGAGAGAGACCAACCCCGCGCAAGACAGAGAGCAAGAGACCGCGAGACCTGACAGC 1440
    |||
Db 1387 TTGAGAGAGACCAACCCCGCGCAAGACAGAGAGCAAGAGACCGCGAGACCTGACAGC 1446
QY 1441 CTGAAGAGCCTGTTGGCAACGACCCCTGAGCCAGTAA 1479
    |||
Db 1447 CTGAAGAGCCTGTTGGCAACGACCCCTGAGCCAGAAA 1485
```

```
RESULT 9
ADCl3257
ID ADCl3257 standard; DNA; 4615 BP.
```

```
XX
AC ADCl3257;
XX
DT 18-DEC-2003 (first entry)
XX
```

```
DE DNA of HIV construct gp160mod-1V1-dV2-gagmod-BW965 SEQ ID NO 36.
XX
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
XX Human immunodeficiency virus.
XX
XX WO2003004620-A2.
XX
XX 16-JAN-2003.
XX
XX 05-JUL-2002; 2002WO-US021420.
XX
XX 05-JUL-2001; 2001US-0303192P.
XX
XX 31-AUG-2001; 2001US-0316860P.
XX
XX 16-JAN-2002; 2002US-0349871P.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (UYST-) UNIV STELLENBOSCH.
XX
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ,
XX
XX WPI, 2003-221593/21.
XX
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX
XX Prot, or Rev polypeptide, useful for immunization, or generating
XX
XX packaging cell lines.
XX
XX Disclosure; Fig 33; 301pp; English.
XX
XX
XX The invention relates to a novel expression cassette comprising a
XX
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX
XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX
XX expression cassette can be used to treat HIV type C by gene therapy or
XX
XX used in the development of a vaccine. The gene delivery vector is
XX
XX administered intramuscularly, intramuscularly, intranasally,
XX
XX subcutaneously, intradermally, transdermally, intravenously,
XX
XX intrarectally, orally or intravenously. The expression cassette is useful
XX
XX for immunisation, generating packaging cell lines and producing HIV
XX
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX
XX Type C related sequence of the invention.
XX
XX
XX Sequence 4615 BP; 1030 A; 1587 C; 1354 G; 644 T; 0 U; 0 Other;
SQ
Query Match 99.9%; Score 1477.4; DB 10; Length 4615;
Best Local Similarity 99.9%; Pred. No. 4.3e-185;
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGGGCGCCCGCCAGCATCTCGCGCGCGCAAGCTGACGCCCTGGAGCGCATCCGC 60
    |||
Db 3137 ATGGGGCGCCCGCCAGCATCTCGCGCGCGCAAGCTGACGCCCTGGAGCGCATCCGC 3196
QY 61 CTGCGCCCCCGCGCAAGAGTGTACATGATGAAGCAACCTGTGTGGGCCAGCCGCGAG 120
    |||
Db 3197 CTGCGCCCCCGCGCAAGAGTGTACATGATGAAGCAACCTGTGTGGGCCAGCCGCGAG 3256
QY 121 CTGGAAGAGTTGCCCCCTGAACCCCGCGCTGTGAGAGCAAGGAGGCTGCAAGCAGATC 180
    |||
Db 3257 CTGGAAGAGTTGCCCCCTGAACCCCGCGCTGTGAGAGCAAGGAGGCTGCAAGCAGATC 3316
QY 181 ATCCGCGAGCTGACACCCCGCGCTGAGACCGGCAAGAGAGCTGAAGAGCCTGTTCAAC 240
    |||
Db 3317 ATCCGCGAGCTGACACCCCGCGCTGAGACCGGCAAGAGAGCTGAAGAGCCTGTTCAAC 3376
QY 241 ACCGTGGCCACCTGTACTGTGCTGACAGAGAGATGAGGTCCGACACCAAGAGAGCC 300
    |||
Db 3377 ACCGTGGCCACCTGTACTGTGCTGACAGAGAGATGAGGTCCGACACCAAGAGAGCC 3436
QY 301 CTGACAAGATCGAGAGAGAGCAAGCAAGTGCAGCAGAGATCCAGCAGGCGGAGGCC 360
    |||
Db 3437 CTGACAAGATCGAGAGAGAGCAAGCAAGTGCAGCAGAGATCCAGCAGGCGGAGGCC 3496
```

```
QY 361 GCCGACAAAGGCGCAAGGTGAGCCAGAACTAACCCCATCGTGCAGAAACCTGCAAGGCGCAGATG 420
    |||
Db 3497 GCCGACAAAGGCGCAAGGTGAGCCAGAACTAACCCCATCGTGCAGAAACCTGCAAGGCGCAGATG 3556
QY 421 GTGCACCAAGGCCATCAGCCCCCGCACCTTGAAAGCGCTGGGGTGAAAGGTGATCGAGGAGAAG 480
    |||
Db 3557 GTGCACCAAGGCCATCAGCCCCCGCACCTTGAAAGCGCTGGGGTGAAAGGTGATCGAGGAGAAG 3616
QY 481 GCCTTGAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540
    |||
Db 3617 GCCTTGAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 3676
QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCCACCAAGGCCCGCATGAGATGCTGAAG 600
    |||
Db 3677 GACCTGAACACGATGTTGAACACCGTGGGCGGCCACCAAGGCCCGCATGAGATGCTGAAG 3736
QY 601 GACACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCGTGACACCCCGTGACCGCGGCGCC 660
    |||
Db 3737 GACACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCGTGACACCCCGTGACCGCGGCGCC 3796
QY 661 ATGCCCCCGCGCGCAGATGCGCGAGCCCGCGCGGACGACATCGCCGCGCACCAAGCAGACC 720
    |||
Db 3797 ATGCCCCCGCGCGCAGATGCGCGAGCCCGCGCGGACGACATCGCCGCGCACCAAGCAGACC 3856
QY 721 CTGCAGGAGCAGATCGCCTGATGACAGCAACCCCCCATCCCGTGGGCGACATCTAC 780
    |||
Db 3857 CTGCAGGAGCAGATCGCCTGATGACAGCAACCCCCCATCCCGTGGGCGACATCTAC 3916
QY 781 AAGCGGTGATCATCTGGGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAGATC 840
    |||
Db 3917 AAGCGGTGATCATCTGGGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAGATC 3976
QY 841 CTGCATCATCAAGCAGGCCCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAG 900
    |||
Db 3977 CTGCATCATCAAGCAGGCCCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAG 4036
QY 901 ACCCTGCGCGCGAGCAGACCCCAAGAGGTGAAGAAGTGAATGACGACACCCCTGCTG 960
    |||
Db 4037 ACCCTGCGCGCGAGCAGACCCCAAGAGGTGAAGAAGTGAATGACGACACCCCTGCTG 4096
QY 961 GTGCAGAAAGCCCAACCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCCGCGCGCAGC 1020
    |||
Db 4097 GTGCAGAAAGCCCAACCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCCGCGCGCAGC 4156
QY 1021 CTGCAGGAGATGATGACCGCCTGCAAGGCGGTGGGCGGCCCAAGGCGCGGTG 1080
    |||
Db 4157 CTGCAGGAGATGATGACCGCCTGCAAGGCGGTGGGCGGCCCAAGGCGCGGTG 4216
QY 1081 CTGGCCGAGGCGATGAGCCAGGCCAACCAAGCGTATGATGACAGAGAACCACTTCAAG 1140
    |||
Db 4217 CTGGCCGAGGCGATGAGCCAGGCCAACCAAGCGTATGATGACAGAGAACCACTTCAAG 4276
QY 1141 GGCCCCCGCGCATGCTCAAGTGTCTCAACTGCGGCAAGAGGGCCCATCGCCCGCAAC 1200
    |||
Db 4277 GGCCCCCGCGCATGCTCAAGTGTCTCAACTGCGGCAAGAGGGCCCATCGCCCGCAAC 4336
QY 1201 TGCCGCGCGCGCGCAAGAGGGCTGTGAAGTGGGCAAGAGGGCCACAGATGAAG 1260
    |||
Db 4337 TGCCGCGCGCGCGCAAGAGGGCTGTGAAGTGGGCAAGAGGGCCACAGATGAAG 4396
QY 1261 GACTGCACCGAGCGCGCAAGCTTCTTGGGCAAGATCTGCGGCAAGAGGGCGC 1320
    |||
Db 4397 GACTGCACCGAGCGCGCAAGCTTCTTGGGCAAGATCTGCGGCAAGAGGGCGC 4456
QY 1321 CCGGCAACTTCTTGAGAGCGCGCGAGCCACCGCGCGCGCGAGAGCTTCCG 1380
    |||
Db 4457 CCGGCAACTTCTTGAGAGCGCGCGAGCCACCGCGCGCGCGAGAGCTTCCG 4516
QY 1381 TTGAGAGAGACCAACCCCGCGCAGAGCAGAGCAAGAGACCGCGAGACCTGACAGC 1440
    |||
Db 4517 TTGAGAGAGACCAACCCCGCGCAGAGCAGAGCAAGAGACCGCGAGACCTGACAGC 4576
QY 1441 CTGAAGAGCCTGTTGCGCAAGACCCCTGAGCAATA 1479
```

```
Db 4577 CTGAAGAGCCTGTTGCGCAAGACCCCTGAGCCAATA 4615
    |||
RESULT 10
ADCL3259
ID ADCL3259 standard; DNA; 4702 BP.
XX
AC ADCL3259;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct gp160mod-TV1-gagmod-BW965 SEQ ID NO 38.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR ) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
XX WPI; 2003-221593/21.
XX
DR
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 35; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramuscularly, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 4702 BP; 1057 A; 1617 C; 1371 G; 657 T; 0 U; 0 Other;

Query Match 99.9%; Score 1477.4; DB 10; Length 4702;
Best Local Similarity 99.9%; Pred. No. 4.3e-185;
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60
    |||
Db 3224 ATGGGCGCGCGCGCAGCATCTGCGCGGCGCAAGCTGAGCGCTGGAGCGCATCCGC 3283
QY 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACCTGTGTGGCCAGCGCGAG 120
    |||
Db 3284 CTGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGCACCTGTGTGGCCAGCGCGAG 3343
QY 121 CTGAGAAAGTTGCGCCTGAACCCCGCGCTGTGAGAGCAGCGAGGCTGCAAGCAGATC 180
    |||
Db 3344 CTGAGAAAGTTGCGCCTGAACCCCGCGCTGTGAGAGCAGCGAGGCTGCAAGCAGATC 3403
```

QY 181 ATCCGCCAGCTGCACCCCGCCCTGACAGCCGGCAGGAGCTGAAGACCTGTTCAC 240  
Db 3404 ATCCGCCAGCTGCACCCCGCCCTGACAGCCGGCAGGAGCTGAAGACCTGTTCAC 3463  
QY 241 ACCGTGGCCACCTGTACTGCGTGACAGAAAGATCGAGGTCCGCGACACCAAGAGGCC 300  
Db 3464 ACCGTGGCCACCTGTACTGCGTGACAGAAAGATCGAGGTCCGCGACACCAAGAGGCC 3523  
QY 301 CTGACAGAATCGAGGAGAGCAGAACTAGTCCAGCAGAAAGATCCAGAGGCCGAGGCC 360  
Db 3524 CTGACAGAATCGAGGAGAGCAGAACTAGTCCAGCAGAAAGATCCAGAGGCCGAGGCC 3583  
QY 361 GCCGACAGGGGCAAGGTGAGCCAGAACTAACCCTGTCAGAACTTCAGGGCCAGATG 420  
Db 3584 GCCGACAGGGGCAAGGTGAGCCAGAACTAACCCTGTCAGAACTTCAGGGCCAGATG 3643  
QY 421 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAAGCTTGGGTGAAGGTGATCGAGAGAA 480  
Db 3644 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAAGCTTGGGTGAAGGTGATCGAGAGAA 3703  
QY 481 GCCTTCAGCCCCCGGAGGTGATCCCGCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCGCAG 540  
Db 3704 GCCTTCAGCCCCCGGAGGTGATCCCGCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCGCAG 3763  
QY 541 GACCTGAACACGATGTGTAACAACCCGTGGGGCGGGCCACCAAGCCGCAATGCAGATGCTGAAG 600  
Db 3764 GACCTGAACACGATGTGTAACAACCCGTGGGGCGGGCCACCAAGCCGCAATGCAGATGCTGAAG 3823  
QY 601 GACACCATCAACGAGAGAGCGCCCGCGAGTGGGACCCGCGTGACCCCGTGACGCGCGGCC 660  
Db 3824 GACACCATCAACGAGAGAGCGCCCGCGAGTGGGACCCGCGTGACCCCGTGACGCGCGGCC 3883  
QY 661 ATCGCCCCCGGCGCAGATGCGCGAGCGCCCGCGGCGAGCGACATCGCGGGCACCACCAAGCAAC 720  
Db 3884 ATCGCCCCCGGCGCAGATGCGCGAGCGCCCGCGGCGAGCGACATCGCGGGCACCACCAAGCAAC 3943  
QY 721 CTGCAGGAGCAGATGCGCTTGATGACCAAGCAACCCCGCCATCCCGGTGGGGCAGATCTAC 780  
Db 3944 CTGCAGGAGCAGATGCGCTTGATGACCAAGCAACCCCGCCATCCCGGTGGGGCAGATCTAC 4003  
QY 781 AAGCGGTGATCATCTCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 840  
Db 4004 AAGCGGTGATCATCTCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 4063  
QY 841 CTGCACATCAAGCAGGGGGCCCCCAAGAGCCCTTCCGCGACTACGTGGAACCGCTTCTCAAG 900  
Db 4064 CTGCACATCAAGCAGGGGGCCCCCAAGAGCCCTTCCGCGACTACGTGGAACCGCTTCTCAAG 4123  
QY 901 ACCCTGCGCGCGCAGCAGAGCACCAGAGGTGAAGAACTGATGACCGACACCTGCTG 960  
Db 4124 ACCCTGCGCGCGCAGCAGAGCACCAGAGGTGAAGAACTGATGACCGACACCTGCTG 4183  
QY 961 GTGCAGAACGCGCAACCCCGCATGCAAGACCATCTGCGCGCTTCTGGCCCCCGCGCCAGC 1020  
Db 4184 GTGCAGAACGCGCAACCCCGCATGCAAGACCATCTGCGCGCTTCTGGCCCCCGCGCCAGC 4243  
QY 1021 CTGAGAGAGATGATGACCGCGCTGCGCAGGGCGGTGGGGCCCCCAAGCCCAAGCGCGCGTG 1080  
Db 4244 CTGAGAGAGATGATGACCGCGCTGCGCAGGGCGGTGGGGCCCCCAAGCGCGCGTG 4303  
QY 1081 CTGGCCGAGCGGATGAGCCAGGCGCAACACCAAGCGTGATGATGAGAAAGCACTTCAAG 1140  
Db 4304 CTGGCCGAGCGGATGAGCCAGGCGCAACACCAAGCGTGATGATGAGAAAGCACTTCAAG 4363  
QY 1141 GGGCCCCCGCGCATGCTCAAGTGTTCACCTGCGGCAAGAGAGGCCACATCGCGCGCAAC 1200  
Db 4364 GGGCCCCCGCGCATGCTCAAGTGTTCACCTGCGGCAAGAGAGGCCACATCGCGCGCAAC 4423  
QY 1201 TGCCCGCGCCCCCGCAAGAGGGCTGCTGAAGTGGCGCAAGAGGGCCACCAAGATGAAG 1260  
Db 4424 TGCCCGCGCCCCCGCAAGAGGGCTGCTGAAGTGGCGCAAGAGGGCCACCAAGATGAAG 4483  
QY 1261 GACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCCCAGGCCAAGGGCGCG 1320

Db 4484 GACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCCCAGCCACAGAGGGCGC 4543  
QY 1321 CCCGGCAACTTCTCTGACAGAGCGCGCCCGAGCCACCGCCCCCGCGAGAGCTTCGC 1380  
Db 4544 CCCGGCAACTTCTCTGACAGAGCGCGCCCGAGCCACCGCCCCCGCGAGAGCTTCGC 4603  
QY 1381 TTGAGAGAGACCAACCCCGCGCGCAAGAGCAGAGAGCAGAGACCTGACCAAGC 1440  
Db 4604 TTGAGAGAGACCAACCCCGCGCGCAAGAGCAGAGAGCAGAGACCTGACCAAGC 4663  
QY 1441 CTGAAGAGCCTGTTGGGCAACGACCCCTGAGCCCAATA 1479  
Db 4664 CTGAAGAGCCTGTTGGGCAACGACCCCTGAGCCCAATA 4702

RESULT 11  
ACA03522  
ID ACA03522 standard; DNA; 4716 BP.  
XX  
AC ACA03522;  
XX  
DT 22-MAY-2003 (first entry)  
XX  
DE Synthetic DNA encoding immunogenic HIV peptide #5.  
XX  
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX  
OS Synthetic.  
XX  
PN WO2003004657-A1.  
XX  
PD 16-JAN-2003.  
XX  
PF 05-JUL-2002; 2002WO-US021421.  
XX  
PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349728P.  
PR 16-JAN-2002; 2002US-0349793P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Zur Megele J, Barnett SW, Lian Y;  
XX  
DR WPI; 2003-221602/21.  
XX  
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
PT especially humans.  
XX  
PS Example 1; Fig 10; 262pp; English.  
XX  
CC The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide  
XX  
SQ Sequence 4716 BP; 1041 A; 1684 C; 1463 G; 528 T; 0 U; 0 Other;  
Query Match 99.9%; Score 1477.4; DB 8; Length 4716;

Best Local Similarity 99.9%; Pred. No. 4.3e-185;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCAGCATCTCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
Db 7 ATGGGCGCGCGCGCAGCATCTCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 66

QY 61 CTGCGCGCGCGCGCAAGAAGTGCTACATGATGAAGCACTGTGTGGCCAGCCGCGAG 120  
Db 67 CTGCGCGCGCGCGCAAGAAGTGCTACATGATGAAGCACTGTGTGGCCAGCCGCGAG 126

QY 121 CTGAGAAGTTGCGCCTGAACCCCGCGCTGTGAGACGAGCGGCTGCAAGCATC 180  
Db 127 CTGAGAAGTTGCGCCTGAACCCCGCGCTGTGAGACGAGCGGCTGCAAGCATC 186

QY 181 ATCCGCGAGCTGCAACCCCGCGCTGCAACCGGCAAGAGCTGAAGCCTGTTCAAC 240  
Db 187 ATCCGCGAGCTGCAACCCCGCGCTGCAACCGGCAAGAGCTGAAGCCTGTTCAAC 246

QY 241 ACCGTGGCCACCCCTGTACTGCGTGCACAGAGAAGATCGAGTCCGCGACCAAGAGGCC 300  
Db 247 ACCGTGGCCACCCCTGTACTGCGTGCACAGAGAAGATCGAGTCCGCGACCAAGAGGCC 306

QY 301 CTGACAAGATCGAGAGAGCAGAACAGTGCAGAGAAAGATCCAGCAGGCCGAGGCC 360  
Db 307 CTGACAAGATCGAGAGAGCAGAACAGTGCAGAGAAAGATCCAGCAGGCCGAGGCC 366

QY 361 GCCGACAAGGGCAGAGGTGAGCCAGAACTAACCCCATCGTGCAAGAACTTGAGGGCCAGATG 420  
Db 367 GCCGACAAGGGCAGAGGTGAGCCAGAACTAACCCCATCGTGCAAGAACTTGAGGGCCAGATG 426

QY 421 GTGCACCAAGGCCATCAGGCCCCCGCACCCCTGAACGCTGGGTGAAGGTATCGAGAGAAAG 480  
Db 427 GTGCACCAAGGCCATCAGGCCCCCGCACCCCTGAACGCTGGGTGAAGGTATCGAGAGAAAG 486

QY 481 GCCTTGAGCCCCCGAGGGTATCCCATGTTTCAACCGCCTGAGCGAGGGGCGCCACCCCGCAG 540  
Db 487 GCCTTGAGCCCCCGAGGGTATCCCATGTTTCAACCGCCTGAGCGAGGGGCGCCACCCCGCAG 546

QY 541 GACCTGAACACAGATGTTGAACACCGTGGGCGGCCACCAAGGCCCGCATGAGATGCTGAAG 600  
Db 547 GACCTGAACACAGATGTTGAACACCGTGGGCGGCCACCAAGGCCCGCATGAGATGCTGAAG 606

QY 601 GACACCATCAACGAGAGGGCGCGCGAGTGGGACCGCGTGACACCCCGTGCAACGCGCGGCC 660  
Db 607 GACACCATCAACGAGAGGGCGCGCGAGTGGGACCGCGTGACACCCCGTGCAACGCGCGGCC 666

QY 661 ATCGCCCCCGCGCAGATGCGCGAGCCCCCGCGGCAAGCATGCGCGGCAACCAAGCAACC 720  
Db 667 ATCGCCCCCGCGCAGATGCGCGAGCCCCCGCGGCAAGCATGCGCGGCAACCAAGCAACC 726

QY 721 CTGCAAGAGCAGATGCGCTGATGACCAAGCAACCCCGCATCCCGTGGGGGAGCATCTAC 780  
Db 727 CTGCAAGAGCAGATGCGCTGATGACCAAGCAACCCCGCATCCCGTGGGGGAGCATCTAC 786

QY 781 AAGCGGTGATCATCTGGGCTGAACAAGATGTGCGAGTGTACAGCCCGTGAGCATC 840  
Db 787 AAGCGGTGATCATCTGGGCTGAACAAGATGTGCGAGTGTACAGCCCGTGAGCATC 846

QY 841 CTGACATCAAGCAGGCGCCCAAGGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 900  
Db 847 CTGACATCAAGCAGGCGCCCAAGGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 906

QY 901 ACCCTGCGCGCGAGCAGACACCCAGAGGTGAAGAAGTGTGATGACGACACCCCTGCTG 960  
Db 907 ACCCTGCGCGCGAGCAGACACCCAGAGGTGAAGAAGTGTGATGACGACACCCCTGCTG 966

QY 961 GTGCAAGAGCCCAACCCCGACTGAAGACCATCTGCGCGCTCTCGGCCCGCGCGCCAGC 1020  
Db 967 GTGCAAGAGCCCAACCCCGACTGAAGACCATCTGCGCGCTCTCGGCCCGCGCGCCAGC 1026

QY 1021 CTGAGAGATGATGACCGCCTGCGCAGGGCGTGGCGGCCCGCAGCCCAAGGCCCGCGTG 1080  
Db 1027 CTGAGAGATGATGACCGCCTGCGCAGGGCGTGGCGGCCCGCAGCCCAAGGCCCGCGTG 1086

Db 1027 CTGAGAGATGATGACCGCCTGCGCAGGGCGTGGCGGCCCGCAGCCCAAGGCCCGCGTG 1086

QY 1081 CTGCGCGAGCGGATGAGCCAGGCGCAACCAACGCTGATGATGACAGAAAGCAACTTCAAG 1140  
Db 1087 CTGCGCGAGCGGATGAGCCAGGCGCAACCAACGCTGATGATGACAGAAAGCAACTTCAAG 1146

QY 1141 GGGCGCGCGCGCATGTCAGTGTCTTCAACTGCGGCAAGAGGGCCCATGCGCCGCAAC 1200  
Db 1147 GGGCGCGCGCGCATGTCAGTGTCTTCAACTGCGGCAAGAGGGCCCATGCGCCGCAAC 1206

QY 1201 TGCGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACAGATGAAG 1260  
Db 1207 TGCGCGCGCGCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACAGATGAAG 1266

QY 1261 GACTGACCGAGCGCGCCAACTTCTGCGCAAGATCTGCGCCAGCCCAAGGGCGGC 1320  
Db 1267 GACTGACCGAGCGCGCCAACTTCTGCGCAAGATCTGCGCCAGCCCAAGGGCGGC 1326

QY 1321 CCGGCAACTTCTGCGAGAGCGCGCGCGAGCCCAACGCGCGCGCGAGAGCTTCGC 1380  
Db 1327 CCGGCAACTTCTGCGAGAGCGCGCGCGAGCCCAACGCGCGCGCGAGAGCTTCGC 1386

QY 1381 TTGAGAGAGACCAACCCCGCGCAAGAGCAGAGAGCAAGAGACCGCGAGACCTGACCAAC 1440  
Db 1387 TTGAGAGAGACCAACCCCGCGCAAGAGCAGAGAGCAAGAGACCGCGAGACCTGACCAAC 1446

QY 1441 CTGAAGAGCCTGTTGCGCAACGAGCCCTGAGCCAGTAA 1479  
Db 1447 CTGAAGAGCCTGTTGCGCAACGAGCCCTGAGCCAGTAA 1485

RESULT 12  
ADCl3238  
ID ADCl3238 standard; DNA, 4716 BP.  
XX  
AC ADCl3238;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA of HIV construct GagProtiNarMultatRevNef\_C SEQ ID NO 17.  
XX  
KW expression cassette; HIV Gag; Env; Int; Nef; p15NaseH; Pol; Tat; Prot;  
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
OS Human immunodeficiency virus.  
XX  
PN WO2003004620-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 05-JUL-2002; 2002WO-US021420.  
XX  
PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX WPI; 2003-221593/21.  
DR  
XX  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15NaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX  
PS Disclosure; Fig 14; 301pp; English.  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15NaseH, Pol, Tat, Prot, or Rev polypeptide. The novel

CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intradermally, intranasally,  
CC subcutaneously, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.

XX  
SQ Sequence 4716 BP; 1041 A; 1684 C; 1463 G; 528 T; 0 U; 0 Other;

Query Match 99.9%; Score 1477.4; DB 10; Length 4716;  
Best Local Similarity 99.9%; Pred. No. 4.3e-185;

Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGCGCCCGCGCCACATCTCTGCGGGCGGCAAGCTGGACGCGCTGGAGCGCATCCGC 60
    |||
Db 7 ATGGGCGCCCGCGCCACATCTCTGCGGGCGGCAAGCTGGACGCGCTGGAGCGCATCCGC 66
QY 61 CTGCGCGCCCGCGGCAAGAAGTGTACATGATGAAGCACTGTGTGTGGCCAGCGCGAG 120
    |||
Db 67 CTGCGCGCCCGCGGCAAGAAGTGTCTATGATGAAGCACTGTGTGTGGCCAGCGCGAG 126
QY 121 CTGAGAGAAGTTCGCGCTGAACCCCGGCTGTGAGACCAAGCGGCTGCAAGCAGATC 180
    |||
Db 127 CTGAGAGAAGTTCGCGCTGAACCCCGGCTGTGAGACCAAGCGGCTGCAAGCAGATC 186
QY 181 ATCCGCCAGCTGCACCCCGCTGACAGACCGGCGAGAGAGCTGAAGAGCTGTTCAC 240
    |||
Db 187 ATCCGCCAGCTGCACCCCGCTGACAGACCGGCGAGAGAGCTGAAGAGCTGTTCAC 246
QY 241 ACCGTGGCCACCTGTACTGCGTGACAGAAAGATCGGTCGCGGACACCAAGAGGCGC 300
    |||
Db 247 ACCGTGGCCACCTGTACTGCGTGACAGAAAGATCGGTCGCGGACACCAAGAGGCGC 306
QY 301 CTGACAAGATCGAGAGAGAGCAACAAGTCCAGCAAGAAGATCCAGAGCGCCAGGCC 360
    |||
Db 307 CTGACAAGATCGAGAGAGAGCAACAAGTCCAGCAAGAAGATCCAGAGCGCCAGGCC 366
QY 361 GCCGACAAGGGCAAGGTGAGCCACAATAACCCCATCTGTGACAACTGCAAGGCCAGATG 420
    |||
Db 367 GCCGACAAGGGCAAGGTGAGCCACAATAACCCCATCTGTGACAACTGCAAGGCCAGATG 426
QY 421 GTGACACGAGCCATCAAGCCCCCGCACCTGAACGCGCTGGGTGAAGGTGATCGAGAGAA 480
    |||
Db 427 GTGACACGAGCCATCAAGCCCCCGCACCTGAACGCGCTGGGTGAAGGTGATCGAGAGAA 486
QY 481 GCGTTCAAGCCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCCACCCCCAG 540
    |||
Db 487 GCGTTCAAGCCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGCGCCACCCCCAG 546
QY 541 GACCTGAACACGATGTGAACACCGGTGGCGGCCACAGGCCCGCATGCAAGTCTGAAG 600
    |||
Db 547 GACCTGAACACGATGTGAACACCGGTGGCGGCCACAGGCCCGCATGCAAGTCTGAAG 606
QY 601 GACACCATCAACGAGAGAGGCGCGCGAGTGGAGCCGCTGCAACCCCGTGCACGCGCGCC 660
    |||
Db 607 GACACCATCAACGAGAGAGGCGCGCGAGTGGAGCCGCTGCAACCCCGTGCACGCGCGCC 666
QY 661 ATCGCCCCCGGCGAGATGCGCGAGCCCCCGGCGAGGACATCGCGCGGCAACCAAGCAC 720
    |||
Db 667 ATCGCCCCCGGCGAGATGCGCGAGCCCCCGGCGAGGACATCGCGCGGCAACCAAGCAC 726
QY 721 CTGCAAGAGCAAGATCGCTGTGATGACCAAGCAACCCCATCCCGCTGGGCGCATCTAC 780
    |||
Db 727 CTGCAAGAGCAAGATCGCTGTGATGACCAAGCAACCCCATCCCGCTGGGCGCATCTAC 786
QY 781 AAGCGGTGATCATCTGGGCTGAACAAGATGTGCGGATGTAACAGCCCGTGAAGCATC 840
    |||
Db 787 AAGCGGTGATCATCTGGGCTGAACAAGATGTGCGGATGTAACAGCCCGTGAAGCATC 846
QY 841 CTGACATCAAGAGAGGCGCCCAAGAGCGCTTCGCGACTAGTGAACGCGCTTCTCAAG 900
    |||
```

```
Db 847 CTGACATCAAGCAGAGGCGCCCAAGAGCGCTTCGCGACTACGTGAGACCGCTTCTCAAG 906
QY 901 ACCCTGCGCGCGAGCAGAGACCCAGAGAGTGAAGAACTGTGATGACCGACACCTGTCTG 960
    |||
Db 907 ACCCTGCGCGCGAGCAGAGACCCAGAGAGTGAAGAACTGTGATGACCGACACCTGTCTG 966
QY 961 GTGCAAGACGCCAACCCCGACTGCAAGACCACTCTGCGCGCTCTCGGCCCGCGCAGC 1020
    |||
Db 967 GTGCAAGACGCCAACCCCGACTGCAAGACCACTCTGCGCGCTCTCGGCCCGCGCAGC 1026
QY 1021 CTGAGAGAGATGATGACCGCTGCGCAGGCGCGTGGCGCGCCAGCCACAAGCGCGCGTG 1080
    |||
Db 1027 CTGAGAGAGATGATGACCGCTGCGCAGGCGCGTGGCGCGCCAGCCACAAGCGCGCGTG 1086
QY 1081 CTGCGCCGAGCGATGAGCCAGGCAACACCAAGCGTGTGATGACGAAGAGCACTTCAAG 1140
    |||
Db 1087 CTGCGCCGAGCGATGAGCCAGGCAACACCAAGCGTGTGATGATGACGAAGAGCACTTCAAG 1146
QY 1141 GCGCCCCGCGCGATCGTCAAGTCTTCAACTGCGGCAAGAGGCGCACATGCGCGCAAC 1200
    |||
Db 1147 GCGCCCCGCGCGATCGTCAAGTCTTCAACTGCGGCAAGAGGCGCACATGCGCGCAAC 1206
QY 1201 TGCGCGCGCGCGCGCAAGAGGGCTGTGAAAGTGCGGCAAGAGGCGCACAGATGAAG 1260
    |||
Db 1207 TGCGCGCGCGCGCGCAAGAGGGCTGTGAAAGTGCGGCAAGAGGCGCACAGATGAAG 1266
QY 1261 GACTGCACCGAGCGCGCAAGCACTTCTGCGGCAAGATCTGCGCCAGCAAGAGGCGCG 1320
    |||
Db 1267 GACTGCACCGAGCGCGCAAGCACTTCTGCGGCAAGATCTGCGCCAGCAAGAGGCGCG 1326
QY 1321 CCGGCAACTTCTTGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
    |||
Db 1327 CCGGCAACTTCTTGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386
QY 1381 TTGAGAGAGACCAACCCCGCGCAGAAGCAGAGAGCAAGAGCGGAGACCTGACAGC 1440
    |||
Db 1387 TTGAGAGAGACCAACCCCGCGCAGAAGCAGAGAGCAAGAGCGGAGACCTGACAGC 1446
QY 1441 CTGAAGAGCTGTTCGCAAGCAAGCCCTGAGCCAGTAA 1479
    |||
Db 1447 CTGAAGAGCTGTTCGCAAGCAAGCCCTGAGCCAGTAA 1485

RESULT 13
ACA03524
ID ACA03524 standard; DNA; 2742 BP.
XX
AC ACA03524;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA encoding immunogenic HIV peptide #7.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002MO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR ) CHIRON CORP.
```

PI Zur Megede J, Barnett SW, Lian Y;  
XX  
DR WPI, 2003-221602/21.  
XX  
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
PT especially humans.  
XX  
PS Example 1; Fig 12; 262pp; English.  
XX  
CC The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide  
XX  
SQ Sequence 2742 BP; 578 A; 1020 C; 859 G; 285 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1475.8; DB 8; Length 2742;  
Best Local Similarity 99.9%; Pred.No.7.5e-185;  
Matches 1477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATGGGCGCCGCGCCAGCATCTGCGCGGCGGCAAGCTGACGCGCTGGAGCGCATCCGC 60  
Db 7 ATGGGCGCCGCGCCAGCATCTGCGCGGCGGCAAGCTGACGCGCTGGAGCGCATCCGC 66  
QY 61 CTGGCGCCCGCGCGCAAGAAGTGTACTATGATGAAGCACTGTGTGGCCAGCCGCGAG 120  
Db 67 CTGGCGCCCGCGCGCAAGAAGTGTACTATGATGAAGCACTGTGTGGCCAGCCGCGAG 126  
QY 121 CTGGAAGAAGTTCGCCCTGAACCCCGGCTGTGAGACGAGGAGCTGAAGACCTGTTCAAC 180  
Db 127 CTGGAAGAAGTTCGCCCTGAACCCCGGCTGTGAGACGAGGAGCTGAAGACCTGTTCAAC 186  
QY 181 ATCCGCCAGCTGACCCCGGCTGTGAGACGAGGAGGAGCTGAAGACCTGTTCAAC 240  
Db 187 ATCCGCCAGCTGACCCCGGCTGTGAGACGAGGAGGAGCTGAAGACCTGTTCAAC 246  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGTCCGCAACCAAGAGGCC 300  
Db 247 ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGTCCGCAACCAAGAGGCC 306  
QY 301 CTGGAACAAGATCGAGAAGAGGAGCAACAAGTGCACGAGAAGATCCAGCAGGCCGAGGCC 360  
Db 307 CTGGAACAAGATCGAGAAGAGGAGCAACAAGTGCACGAGAAGATCCAGCAGGCCGAGGCC 366  
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACTAACCCTATGTCAGAACCTGCAGGGCCAGATG 420  
Db 367 GCCGACAAGGGCAAGGTGAGCCAGAACTAACCCTATGTCAGAACCTGCAGGGCCAGATG 426  
QY 421 GTGACACGAGCCATCAGCCCCCGGCACTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480  
Db 427 GTGACACGAGCCATCAGCCCCCGGCACTGAACGCTGGGTGAAGGTGATCGAGAGAAG 486  
QY 481 GCGTTCAAGCCCGGAGGTATCCCACTGTTCACCGCCCTGAGCGAGGGCCCAACCCCCAG 540  
Db 487 GCGTTCAAGCCCGGAGGTATCCCACTGTTCACCGCCCTGAGCGAGGGCCCAACCCCCAG 546  
QY 541 GACCTGAACACGATGTGAACAACCGTGGGGCGGCAACGAGCCGCAATGCAGATGCTGAAG 600  
Db 547 GACCTGAACACGATGTGAACAACCGTGGGGCGGCAACGAGCCGCAATGCAGATGCTGAAG 606  
QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCCGCCGCC 660  
Db 607 GACACCATCAACGAGAGAGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCCGCCGCC 666

QY 661 ATGCCCCCGGCGAGATGCGGAGAGCCCGGCGGCGAGCATGCGCGGCAACCAAGCAGACC 720  
Db 667 ATGCCCCCGGCGAGATGCGGAGAGCCCGGCGGCGAGCATGCGCGGCAACCAAGCAGACC 726  
QY 721 CTGAGAGAGAGATCGCCTGTGATGACCAAGCAACCCCGCATCCCGTGGGCGACATCTAC 780  
Db 727 CTGAGAGAGAGATCGCCTGTGATGACCAAGCAACCCCGCATCCCGTGGGCGACATCTAC 786  
QY 781 AAGCGGTGATCATCTGGGCTGTGAACAAGATCGTGGATGTACAGCCCGTGAACATC 840  
Db 787 AAGCGGTGATCATCTGGGCTGTGAACAAGATCGTGGATGTACAGCCCGTGAACATC 846  
QY 841 CTGACATCAAGCAGGGGCGGCAAGAGAGCCCTTCGCGACTACGTGAGCCGCTTCTCAAG 900  
Db 847 CTGACATCAAGCAGGGGCGGCAAGAGAGCCCTTCGCGACTACGTGAGCCGCTTCTCAAG 906  
QY 901 ACCCTGCGCGCGGAGCAGAGCAACCCAGAGAGTGAAGACTGTGATGACCGACACCTGCTG 960  
Db 907 ACCCTGCGCGCGGAGCAGAGCAACCCAGAGAGTGAAGACTGTGATGACCGACACCTGCTG 966  
QY 961 GTGAGAACGCGCAACCCCGCATGCAAGACCATCTGCGCGCTTCGGGCGCGCGCAGC 1020  
Db 967 GTGAGAACGCGCAACCCCGCATGCAAGACCATCTGCGCGCTTCGGGCGCGCGCAGC 1026  
QY 1021 CTGAGAGAGATGATGACCGGCTGCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1080  
Db 1027 CTGAGAGAGATGATGACCGGCTGCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1086  
QY 1081 CTGCGCGAGGCGGATGAGCCAGGCGCAACCAAGCGTGTATGTCAGAAGAGCACTTCAAG 1140  
Db 1087 CTGCGCGAGGCGGATGAGCCAGGCGCAACCAAGCGTGTATGTCAGAAGAGCACTTCAAG 1146  
QY 1141 GGCCCCCGGCGCATCTGCAAGTCTTCAACTGCGGCAAGGAGGCGCACATCGCGCGCAAC 1200  
Db 1147 GGCCCCCGGCGCATCTGCAAGTCTTCAACTGCGGCAAGGAGGCGCACATCGCGCGCAAC 1206  
QY 1201 TGCCTGCGGCGGCGGCAAGAGGCGTGTGAAAGTGCAGGAGGCGGCAAGAGATGAAG 1260  
Db 1207 TGCCTGCGGCGGCGGCAAGAGGCGTGTGAAAGTGCAGGAGGCGGCAAGAGATGAAG 1266  
QY 1261 GACTGACCGAGGCGGCAAGGCGCAACTTCTGCGGCAAGATCTGCGGCAAGGAGGCGG 1320  
Db 1267 GACTGACCGAGGCGGCAAGGCGCAACTTCTGCGGCAAGATCTGCGGCAAGGAGGCGG 1326  
QY 1321 CCGGCGCACTTCTGTCAGAGGCGGCGGAGCGGCGGCGGCGGCGGCGGAGCTTCGC 1380  
Db 1327 CCGGCGCACTTCTGTCAGAGGCGGCGGAGCGGCGGCGGCGGCGGCGGAGCTTCGC 1386  
QY 1381 TTCAGAGAGACCAACCCCGGCGGCAAGAGAGCAAGGAGCGGAGACCTGAACAGC 1440  
Db 1387 TTCAGAGAGACCAACCCCGGCGGCAAGAGAGCAAGGAGCGGAGACCTGAACAGC 1446  
QY 1441 CTGAAGAGCTGTTCGCAACGACCCCTGAGCGCAATA 1479  
Db 1447 CTGAAGAGCTGTTCGCAACGACCCCTGAGCGCAATA 1485  
RESULT 14  
AD13241  
ID AD13241 standard; DNA; 2742 BP.  
XX  
XX AD13241;  
AC AD13241;  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA of HIV construct GagTatRevNef\_C SEQ ID NO 20.  
XX  
XX expression cassette; HIV Gag; Env; Int; Nef; p15RaseH; Pol; Tat; Prot;  
KM Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX  
OS Human immunodeficiency virus.  
XX

PN WO2003004620-A2.  
XX 16-JAN-2003.  
PD  
XX  
XX 05-JUL-2002; 2002WO-US021420.  
PF  
XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX WPI, 2003-221593/21.  
DR  
XX  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX  
XX Disclosure; Fig 17; 301bp; English.  
PS  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramusosally, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX  
XX Sequence 2742 BP; 578 A; 1020 C; 859 G; 285 T; 0 U; 0 Other;  
SQ  
Query Match 99.8%; Score 1475.8; DB 10; Length 2742;  
Best Local Similarity 99.9%; Pred. No. 7.5e-185;  
Matches 1477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATGGGCGCGCGCGCGCATCTCTGCGCGCGCGCGCAAGCTGACCGCTGGAGCGCATCCGC 60  
DB 7 ATGGGCGCGCGCGCGCGCATCTCTGCGCGCGCGCGCAAGCTGACCGCTGGAGCGCATCCGC 66  
QY 61 CTGCGCGCGCGCGCGCGCAAGAGTGTCTACATGATGAGCAACCTGTGTGGCCAGCGCGGAG 120  
DB 67 CTGCGCGCGCGCGCGCGCAAGAGTGTCTACATGATGAGCAACCTGTGTGGCCAGCGCGGAG 126  
QY 121 CTGGAAGAAGTTGCGCCCTGAACCCCGGCTGTGTGAGACCAAGGAGGCTGCAAGCAGATC 180  
DB 127 CTGGAAGAAGTTGCGCCCTGAACCCCGGCTGTGTGAGACCAAGGAGGCTGCAAGCAGATC 186  
QY 181 ATCCGCCAGCTGCAACCCCGCTGCAAGCCGCGAGCGAGGAGCTGAAGAGCCTGTTCAAC 240  
DB 187 ATCCGCCAGCTGCAACCCCGCTGCAAGCCGCGAGCGAGGAGCTGAAGAGCCTGTTCAAC 246  
QY 241 ACCGTGGCCACCTGTACTGCTGTGCAAGAAAGATCGAGGTCGCGACACCAAGGAGGCC 300  
DB 247 ACCGTGGCCACCTGTACTGCTGTGCAAGAAAGATCGAGGTCGCGACACCAAGGAGGCC 306  
QY 301 CTGGAACAAGATCGAGAGGAGAGCAACAAGTGCAGCAGAAAGATCCAGCAGGCCGAGGCC 360  
DB 307 CTGGAACAAGATCGAGAGGAGAGCAACAAGTGCAGCAGAAAGATCCAGCAGGCCGAGGCC 366  
QY 361 GCCGACAAGGGCAAGGTGAGCGCAAGACTACCCCATCGTGCAAACTGCAAGGCCAGATG 420  
DB 367 GCCGACAAGGGCAAGGTGAGCGCAAGACTACCCCATCGTGCAAACTGCAAGGCCAGATG 426  
QY 421 GTGCACCAAGGCGATCAGCCCGCGCAACCTGAACGCTGGGTGAAGGTGATGAGAGGAAG 480  
DB 427 GTGCACCAAGGCGATCAGCCCGCGCAACCTGAACGCTGGGTGAAGGTGATGAGAGGAAG 486

QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
DB 487 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 546  
QY 541 GACCTGAACAGATGTTGAACAACCGTGGCGGCCACCAAGGCCCATGCAAGATGCTGAAG 600  
DB 547 GACCTGAACAGATGTTGAACAACCGTGGCGGCCACCAAGGCCCATGCAAGATGCTGAAG 606  
QY 601 GACACCATCAACGAGAGGGCGCCGAGTGGGACCGCGTGCAACCCCGTGCAACCGCGGCC 660  
DB 607 GACACCATCAACGAGAGGGCGCCGAGTGGGACCGCGTGCAACCCCGTGCAACCGCGGCC 666  
QY 661 ATCGCCCCCGGCGAGATCGCGGAGCGCCCGCGGACGACATCGCGGACCAACGACACC 720  
DB 667 ATCGCCCCCGGCGAGATCGCGGAGCGCCCGCGGACGACATCGCGGACCAACGACACC 726  
QY 721 CTGCAGAGAGATCGCGCTGATGACCAAGCAACCCCCCATCCCGTGCGGCGACATCTAC 780  
DB 727 CTGCAGAGAGATCGCGCTGATGACCAAGCAACCCCCCATCCCGTGCGGCGACATCTAC 786  
QY 781 AAGCGGTGATCATCTCTGGGCTGAACAAGATCGTGCGGATGTACAAGCCCGTGAGCATC 840  
DB 787 AAGCGGTGATCATCTCTGGGCTGAACAAGATCGTGCGGATGTACAAGCCCGTGAGCATC 846  
QY 841 CTGCACATCAAGCAGAGGGCGCCCAAGAGCGCTTCGCGACTACGTGAGCGCTTCTTAAG 900  
DB 847 CTGCACATCAAGCAGAGGGCGCCCAAGAGCGCTTCGCGACTACGTGAGCGCTTCTTAAG 906  
QY 901 ACCCTGCGCGCGGAGCAGAGCAACCAAGAGGTGAAGACTGATGACCGCAACCTGCTG 960  
DB 907 ACCCTGCGCGCGGAGCAGAGCAACCAAGAGGTGAAGACTGATGACCGCAACCTGCTG 966  
QY 961 GTGCAGAACGCCAACCCCGCACTGCAAGACCACTCTCGCGCTCTCGCGCGCGCGCGCAGC 1020  
DB 967 GTGCAGAACGCCAACCCCGCACTGCAAGACCACTCTCGCGCTCTCGCGCGCGCGCGCAGC 1026  
QY 1021 CTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCAAGAGCGCGCGCTG 1080  
DB 1027 CTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCAAGAGCGCGCGCTG 1086  
QY 1081 CTGGCGGAGCGATGAGCGGCGCAACCAAGCGTGATGATGCAAGAGCAACTTCAAG 1140  
DB 1087 CTGGCGGAGCGATGAGCGGCGCAACCAAGCGTGATGATGCAAGAGCAACTTCAAG 1146  
QY 1141 GGGCGCGCGCGCATCTGTCAAGTGTCTCAACTGCGGCAAGAGGGCGCACATCGCCGCAAC 1200  
DB 1147 GGGCGCGCGCGCATCTGTCAAGTGTCTCAACTGCGGCAAGAGGGCGCACATCGCCGCAAC 1206  
QY 1201 TGCCCGCGCGCGCGCAAGAGGGCTGTGGAAGTGTGGCGCAAGAGGGCGCACAGATGAAG 1260  
DB 1207 TGCCCGCGCGCGCGCAAGAGGGCTGTGGAAGTGTGGCGCAAGAGGGCGCACAGATGAAG 1266  
QY 1261 GACTGCACCGAGCGCGCAAGCTTCTGTGGGCAAGATCTGGCCAGGCCCAAGAGGGCGCGC 1320  
DB 1267 GACTGCACCGAGCGCGCAAGCTTCTGTGGGCAAGATCTGGCCAGGCCCAAGAGGGCGCGC 1326  
QY 1321 CCGGCAACTTCTCTGCAAGCGCGCGCGCGCAAGCCGCGCGCGCGCGGAGAGCTTCCGC 1380  
DB 1327 CCGGCAACTTCTCTGCAAGCGCGCGCGCGCAAGCCGCGCGCGCGCGGAGAGCTTCCGC 1386  
QY 1381 TTGAGAGAGCAACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1387 TTGAGAGAGCAACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
QY 1441 CTGAAGAGCTGTGTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
DB 1447 CTGAAGAGCTGTGTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485

RESULT 15  
ADCl3230  
ID ADCl3230 standard; DNA; 3930 BP.

XX AC ADCl3230;  
XX 18-DEC-2003 (first entry)  
XX DNA of HIV construct GagComp1Polmut\_C SEQ ID NO 9.  
DE expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
KM Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX Human immunodeficiency virus.  
XX WO2003004620-A2.  
XX 16-JAN-2003.  
XX 05-JUL-2002; 2002WO-US021420.  
XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
PI WPI; 2003-221593/21.  
XX  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX  
XX Disclosure; Fig 6; 301pp; English.  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramusosally, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX  
XX Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;  
SQ  
Query Match 99.8%; Score 1475.8; DB 10; Length 3930;  
Best Local Similarity 99.9%; Pred. No. 7.1e-185;  
Matches 1477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 301 CTGACCAAGATCGAGGAGGAGCAGAACTAGTCCAGCAGAAAGATCCAGAGCCGAGGCC 360  
Db 307 CTGACCAAGATCGAGGAGGAGCAGAACTAGTCCAGCAGAAAGATCCAGAGCCGAGGCC 366  
QY 361 GCCGACCAAGGCGCAAGGTGAGCCAGAACTAACCCTATCGTGCAAGAACTTCAGAGGCCAGATG 420  
Db 367 GCCGACCAAGGCGCAAGGTGAGCCAGAACTAACCCTATCGTGCAAGAACTTCAGAGGCCAGATG 426  
QY 421 GTGACCAAGGCGCATGACCCCGCAGCCTGGAACGCTGGGTGAAGGTGATCGAGAGAG 480  
Db 427 GTGACCAAGGCGCATGACCCCGCAGCCTGGAACGCTGGGTGAAGGTGATCGAGAGAG 486  
QY 481 GCCTTCAAGCCCGAGGTGATCCCATGTTCAACCGCCTGAGCGAGGGCGCCACCCCCAG 540  
Db 487 GCCTTCAAGCCCGAGGTGATCCCATGTTCAACCGCCTGAGCGAGGGCGCCACCCCCAG 546  
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGCGCCACAGGCCGCAATGCAAGTCTGAAG 600  
Db 547 GACCTGAACACGATGTTGAACACCGTGGCGCGCCACAGGCCGCAATGCAAGTCTGAAG 606  
QY 601 GACACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCAACCCCGTGCAACGCCGCC 660  
Db 607 GACACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCAACCCCGTGCAACGCCGCC 666  
QY 661 ATGCCCCCGGCGCAAGATGCGCGAGAGCCCGCGGCAAGCATCGCCGCGCACCAAGCACC 720  
Db 667 ATGCCCCCGGCGCAAGATGCGCGAGAGCCCGCGGCAAGCATCGCCGCGCACCAAGCACC 726  
QY 721 CTGCAAGAGCAGATCGCCTTGATGACCAAGCAACCCCCCATCCCGGTGGCGACATCTAC 780  
Db 727 CTGCAAGAGCAGATCGCCTTGATGACCAAGCAACCCCCCATCCCGGTGGCGACATCTAC 786  
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 840  
Db 787 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCGTGAGCATC 846  
QY 841 CTGACATCAAGCAGGAGCCCAAGAGACCCCTTCCGCACTACGTGACCGCTTCTTCAAG 900  
Db 847 CTGACATCAAGCAGGAGCCCAAGAGACCCCTTCCGCACTACGTGACCGCTTCTTCAAG 906  
QY 901 ACCCTGCGCGCGGAGAGAGCAACCCAGAGAGGTGAAGAACTGATGACCGACACCTTCTG 960  
Db 907 ACCCTGCGCGCGGAGAGAGCAACCCAGAGAGGTGAAGAACTGATGACCGACACCTTCTG 966  
QY 961 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCGCGCAGC 1020  
Db 967 GTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCGCGCAGC 1026  
QY 1021 CTGAGGAGATGATGACCGCTTGCAGGCGCGTGGCGGCGCCAGCACAAGGCCCGGTG 1080  
Db 1027 CTGAGGAGATGATGACCGCTTGCAGGCGCGTGGCGGCGCCAGCACAAGGCCCGGTG 1086  
QY 1081 CTGGCCGAGGCGATGAGCCAGGCCAACCAACAGCGTGATGATGACAGAGCAACTTCAAG 1140  
Db 1087 CTGGCCGAGGCGATGAGCCAGGCCAACCAACAGCGTGATGATGACAGAGCAACTTCAAG 1146  
QY 1141 GGGCCCCGGCGCATCTGTAAGTGTCTTCAACTGCGGCAAGGAGGCCCATGCGCCGCAAC 1200  
Db 1147 GGGCCCCGGCGCATCTGTAAGTGTCTTCAACTGCGGCAAGGAGGCCCATGCGCCGCAAC 1206  
QY 1201 TGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTCCGCAAGAGGGCCACCAAGATGAAG 1260  
Db 1207 TGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTCCGCAAGAGGGCCACCAAGATGAAG 1266  
QY 1261 GACTGACCGAGCGCGCAAGCTTCTGCGCAAGATCTGCGCCAGCCCAAGGGCGC 1320  
Db 1267 GACTGACCGAGCGCGCAAGCTTCTGCGCAAGATCTGCGCCAGCCCAAGGGCGC 1326  
QY 1321 CCCGCAACTTCTGCAAGAGCGCGCCGAGCCCAAGCCCGCCCGCGAGAGCTTCGC 1380  
Db 1327 CCCGCAACTTCTGCAAGAGCGCGCCGAGCCCAAGCCCGCCCGCGAGAGCTTCGC 1386

QY	1381	TTTCGAGGAGACCA	CCCCCGGCCAGAA	GCAGACGACCGGAGAC	CTTGACCA	G	1440
Db	1387	TTTCGAGGAGACCA	CCCCCGGCCAGAA	GCAGACGACCGGAGAC	CTTGACCA	G	1446
QY	1441	CTGAAGAGCCTGT	TCGGCAACGAC	CCCCCTGAGC	CA	G	1479
Db	1447	CTGAAGAGCCTGT	TCGGCAACGAC	CCCCCTGAGC	CA	G	1485

Search completed: March 12, 2005, 17:35:52  
Job time : 843.981 Secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 16:34:11 ; Search time 5208.18 Seconds  
(without alignments)  
10809.355 Million cell updates/sec

Title: US-09-475-704A-3  
Perfect score: 1479  
Sequence: 1 atggcgccgcgcgcgcacat.....acgacccctgagccagtaa 1479

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_htc:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	82.6	5.6	746	7 CK215205	CK215205 FGAS02715
2	81.2	5.5	749	5 BQ744212	BQ744212 WHE4113 A
3	80	5.4	1064	7 CK163121	CK163121 FGAS01573
4	80	5.4	2277	9 CL968487	CL968487 OSIFCC040
C 5	79.8	5.4	1126	7 CK211108	CK211108 FGAS02294
6	79.4	5.4	826	7 CN128942	CN128942 RHOH1_32
7	78	5.3	807	6 CB673622	CB673622 OSJNE08D
8	78	5.3	821	6 CB662794	CB662794 OSJNE07C
9	78	5.3	824	6 CB666192	CB666192 OSJNEd12P
10	78	5.3	851	6 CB645622	CB645622 OSJNE07H
11	76.6	5.2	807	7 CB673355	CB673355 OSJNE07N
12	76.6	5.2	1138	7 CK163513	CK163513 FGAS01614
13	76.4	5.2	812	6 CB681973	CB681973 OSJNEf08M
14	76.4	5.2	832	6 CB684834	CB684834 OSJNEf14L
15	76.2	5.2	759	2 BF259495	BF259495 HVSMEf001
16	76.2	5.2	1173	9 CL979928	CL979928 OSIFCC045
17	75.6	5.1	562	2 BF484304	BF484304 WHE2321 F
18	75.6	5.1	649	6 CB870888	CB870888 HCl5J04W
19	75.4	5.1	834	6 CB678700	CB678700 OSJNEf01C
20	75.4	5.1	840	6 CB673814	CB673814 OSJNE08I
21	74.8	5.1	766	6 CB629976	CB629976 OSIEB06K
22	74.8	5.1	797	6 CB658304	CB658304 OSJNEc14E
23	74.8	5.1	810	6 CB618374	CB618374 OSIEB02D
24	74.8	5.1	823	6 CB627740	CB627740 OSIEB020

25	74.8	5.1	826	6 CB641789	CB641789 OSJNEb01J
26	74.8	5.1	838	6 CB659169	CB659169 OSJNEc15O
27	74.8	5.1	865	6 CB628660	CB628660 OSIEB04F
28	74.6	5.0	708	6 CD878344	CD878344 AZ04.102I
29	74.6	5.0	818	6 CB652454	CB652454 OSJNEC02G
30	74.4	5.0	1962	9 CL961326	CL961326 OSIFCC005
31	74.2	5.0	785	4 BG369139	BG369139 HVSME1002
32	74.2	5.0	933	8 CC391778	CC391778 PUHJ170TB
33	73.6	5.0	791	6 CB650382	CB650382 OSJNEb14M
34	73.4	5.0	731	5 BQ752847	BQ752847 WHE4119 G
35	73	4.9	2629	9 CL955065	CL955065 OSJRU000
36	72.6	4.9	2598	3 AY103647	AY103647 Zea mays
37	72.4	4.9	648	2 BE517305	BE517305 WHE0616 A
38	72.2	4.9	803	6 CB633391	CB633391 OSIEB12D
39	72.2	4.9	1308	3 AY104577	AY104577 Zea mays
40	72	4.9	1728	9 CL973038	CL973038 OSIFCC042
41	71.8	4.9	764	5 BQ804819	BQ804819 WHE3559 C
42	71.4	4.8	718	6 CD938268	CD938268 OV.109123
43	71.4	4.8	718	6 CD938289	CD938289 OV.109K04
44	71.2	4.8	782	6 CB661708	CB661708 OSJNEd04P
45	71.2	4.8	2517	9 CL972413	CL972413 OSIFCC022

ALIGNMENTS

RESULT 1  
CK215205/c  
LOCUS  
DEFINITION CK215205 746 bp mRNA linear EST 09-DEC-2003  
FGAS027158 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
aestivum cDNA, mRNA sequence.

ACCESSION CK215205  
VERSION CK215205.1 GI:39621309  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

REFERENCE Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 746)

AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Budes,F., Frick,M., Gaudet,D.,  
Genewein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,  
Linke,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilsson,D.,  
Peniket,C., Roach,J.L. and Sarhan,F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops

Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_est@cs.usask.ca

TITLE  
JOURNAL  
COMMENT  
This sequence is the direct result of the Base calling software  
phred (default parameter). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [13,662].  
Place: L6B016 row: K column: 15.

FEATURES  
source

1..746  
Location/Qualifiers  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf, Vector: pCMV.SPORT6, Crown  
(50%) and leaf (50%) tissues from wheat cultivar Norstar  
after short exposure times to low temperature in the light  
and in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20Cs from wheat cultivar Norstar after  
short exposure times to low temperature in the light and

in the dark. 12 mRNA populations were combined before constructing the library. The first 6 populations: After 7 days of growth at 20, wheat plants were transferred to 4C in the light. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. The last 6 populations: After 7 days of growth at 20C, wheat plants were transferred to 4C in the dark. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI. In addition, this library used a primer for second strand synthesis that annealed to an artificial sequence (RNA oligo) added before first strand synthesis. Therefore when sequences from EST generated from this library will be masked for vector and adaptor sequences, an additional masking step will have to be included to mask this RNA oligo that is common to all clones (sequence CGACTGAGACGAGGACACTGACATGAGTGAAGAGTAA)." .

## ORIGIN

Query Match 5.6%; Score 82.6; DB 7; Length 746;  
Best Local Similarity 48.4%; Pred. No. 4.6e-05;  
Matches 288; Conservative 0; Mismatches 301; Indels 6; Gaps 2;

```

QY 124 GAGAACTTGGCCCTGAACCCCGGCTGTGAGACGAGGGCTGCAAGCATCATC 183
    |||||
Db 737 GAGATNNCCGAAAAACACCCATGAACTGATGACGACGTGGCGGACGAGGCCAGTTG 678

QY 184 CGCCAGCTGCACCCCGCCCTGCAGACCGGCGAGGAGCTGAAGAGCCTGTTCAACACC 243
    |||||
Db 677 CTCAACATGCTGCTCAAGCTCATCGGCGCCAGAAAGACCATGAGATCGGCGTTACACC 618

QY 244 GTGGCCACCTGTACTGCGTGCACGAGAAAGATCGAGSTCCGCGACCAAGAGGCCCTG 303
    |||||
Db 617 GGGTACTCCCTCTCGCCACCGCGCTGCCATCCCCGACGACGCGACCATCTTGGCCATG 558

QY 304 GACAAGATCGAGGAGGACGAAACAAGTCCAGCAGAAAGATCCAGCAGCGCCGAGGCC 363
    |||||
Db 557 GACATCAACCGCGAGAACTACGAGCTGGGGTGGCTGTCATGAGAAAGCGCGGTGGCG 498

QY 364 GACAAAGGCGAAGTGAGCCAGAACTACCCCATCTGTGAGAACTGAGGCGCCAGATG 423
    |||||
Db 497 CACAAGATGACTTCCGCGAGGG---CCCGCGCTGCGGTGAGACGCGTCTGGAG 441

QY 424 CACGAGGCCATCAGCCCCCGCACCCCTGAACGCGCTGGGTGAAGTGATCGAGAGAA 483
    |||||
Db 440 GACGAGGCCCAACCAAGGACCTTGACTTCTGTTCTGTGACGCGCCAGCAAGGAACTAC 381

QY 484 TTCAGCCCCCGAGGTGATCCCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCGAG 543
    |||||
Db 380 CTCACACTACCGACGAGCGCCTCATGAAAGCTGTCAGCTCGGCGCCTCTCGGCTACGAC 321

QY 544 CTGAACACGATGTTGAACACCGTGGGGCGGCGCACGAGCGCCCATGTCAGATGCTGAAG 603
    |||||
Db 320 ---AACACGCTTGAAACGGCTTCGTGCTCCCGCGCGCCCGCCCATGCGCAAGTAC 264

QY 604 ACCATCAACGAGGAGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCCATC 663
    |||||
Db 263 ATCCGCTACTACCGGAGCTTCTGCTCTGACCTCAACAAGGCGCTCGCGCGGACGAGCGC 204

QY 664 GCGCCCGGCGAGATGCGCGAGCGCGCGCGGAGCATGCGCGCACCAACCAAGCA 718
    |||||
Db 203 GTGAGATCTGCGCAGCTCCCGCTGGCGGAGCGGATCACCCTTGCGCGCGCGCA 149

```

RESULT 2  
BQ744212 749 bp mRNA linear EST 17-JUL-2002  
LOCUS BQ744212  
DEFINITION WHB4113\_A01 A01ZS wheat salt-stressed root cDNA library Triticum  
aestivum cDNA clone WHB4113\_A01\_A01, mRNA sequence.  
ACCESSION BQ744212

VERSION BQ744212.1 GI:21890999  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.  
1 (bases 1 to 749)

AUTHORS Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J., Iazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.  
TITLE The structure and function of the expressed portion of the wheat genomes - Salt-stressed root cDNA library  
JOURNAL Unpublished (2002)

COMMENT Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818

Email: oanderson@pw.usda.gov  
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: SK primer.

FEATURES  
source Location/Qualifiers

```

1..749
/organism="Triticum aestivum"
/mol type="mRNA"
/cultivar="Chinese Spring"
/db xref="taxon:4565"
/clone="WHB4113_A01_A01"
/tissue type="Roots"
/dev stage="Full tillering"
/lab host="E. coli SOLR"
/clone lib="wheat salt-stressed root cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid
bluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Hydropic
plants grown to full tillering stage were treated with 150
mM NaCl for either 12 hours or 7 days. Root tissues of the
plants subjected to both types of treatment were collected
separately at University of California, Davis (E. Akhunov
and K. Deal in J. Dvorak's Lab). Total RNA was prepared
separately from the two samples (12h and 7day treatments),
and equal amount of RNA was then pooled. PolyA RNA was
purified from the pooled RNA, a cDNA library was made, and
the cDNA clones were in vivo excised to give Bluescript
SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak)
at the University of California, Davis. Colony plating,
plasmid DNA preparations and DNA sequencing were performed
in the OD Anderson lab (all other authors)."

```

## ORIGIN

Query Match 5.5%; Score 81.2; DB 5; Length 749;  
Best Local Similarity 47.5%; Pred. No. 8.4e-05;  
Matches 282; Conservative 0; Mismatches 303; Indels 9; Gaps 1;

```

QY 280 GTCCGCGACCAAGAGGCGCCCTGACAAAGATCGAGGAGGACGAGAAAGTGCCAGCAG 339
    |||||
Db 19 GTGACGCGCAAGCGCGCGCCCAAGCACCCCGCGCTTCGGGCCCAAGTGCGCGCTTCAAC 78

QY 340 AAGATCCAGCAGGCGCGAGGCGCGCGACAAAGGCGAAGTGAGCCAGAACTAACCCCAT 399
    |||||
Db 79 CGGCCCGAGCTCCCGCCAGCGCGCGCGAGGAGGAGTCAAGGCCAAGCGCGCGAGGCGGAG 138

QY 400 CAGAACCTGCAAGGCGCGAGATGTGCACCAAGGCCATCAGCCCCCGCACCTGAACGCGCT 459
    |||||
Db 139 AAGAGCGCGCGCGCGCGCGCGGAGGCGGACCTACCAAGAACCAACCGCTCCACCAACCA 198

QY 460 GTGAAGGTGATCGAGGAGAAAGGCGCTTCAAGCCCCCGAGGTGATCCCATGTTCAACGCG 519
    |||||
Db 199 TTCCGCGGCAACCAAGCAGAAACACCGGCAAGCGCTTCGTCACATCCAGCCCGACTGAGC 258

QY 520 AGCGAGGCGCGCACCCCGCCAGACCTGAACACGATGTTGAACACCGTGGGCGGCGCAG 579
    |||||

```

Db 259 ATCTCGACGACATCCCTTTGCGCACTTCAACCAAGCTCTCTTCCGCCGTCGCCGACGAG 318  
QY 580 GCCCGCATGACAGATGCTGAAGACACCATCAACGAGAGGCCGCCGAGTGGAGCCGCTG 639  
Db 319 CCCGAGGACCTGCTCGTCTGCGGGCGCCGCTGACTCGTACGACCGCGCTTACGACCGCGCTC 378  
QY 640 CACCCCGTCACGCGCCGCCCATCGCCCCCGGCAGATGCGCGAGCCCGCGCGAGCGAC 699  
Db 379 AACCCCAAGACGCGCCCGCGCTCGAGCGCTTCAAGAACCGCCAGTTCTTCAAGATCAC 438  
QY 700 ATGCGCGGACCAACGACGACCTGTC-----AGAGCAGATGCGCTGATGACGAC 750  
Db 439 ACCACGACGACCGCCCATCATCGCGCGCTCGCCGAGAGGACCAAGCCAGCTTTCGCC 498  
QY 751 AACCCCGCATCCCGCGGGGACATCTCAAGCGGTGATCATCTGGGCGCTGAACAAG 810  
Db 499 ACCGAGCCATCTCGCGCGCCCTCATGTGACGCCCCGACGATCTCTCTGGGACATT 558  
QY 811 ATCGTGGGATGTACAGCCCCGTGAGCATCTGACATCAAGCAGGCCCCAAG 864  
Db 559 GTCTGACAGCGCGTGGCAACAGCTTTCTTTCGACAGCGCGAGGGCTCCAG 612

RESULT 3  
CK163121 1064 bp mRNA linear EST 05-DEC-2003  
LOCUS FGAS015739 Triticum aestivum FGAS: Library 4 Gate 8 Triticum  
DEFINITION aestivum cDNA, mRNA sequence.  
ACCESSION CK163121  
VERSION CK163121.1 GI:38993028  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

REFERENCE  
AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D., Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A., Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D., Penniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
CONTACT: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_esteecs.usask.ca

FEATURES  
Source  
1.1064  
Location/Qualifiers  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 4 Gate 8"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6;  
Conditions for growth: Seeds were germinated in a  
water-saturated mix (50% black earth and 50% Promix) in a  
growth chamber for 7 days under an irradiance of 200 mmol  
m-2 sec-1. The temperature was maintained at 20 degrees C  
with a 15-hr photoperiod under a relative humidity of 70%.  
After this period watering of plants was stopped. Four  
time points were sampled during a two week period; the  
first after wilting was observed and the last, two weeks  
later, consisted of live crown and leaf tissue (leaf

tissue that was yellow was not included in sampled material). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

Query Match 5.4%; Score 80; DB 7; Length 1064;  
Best Local Similarity 48.4%; Pred. No. 0.00014;  
Matches 283; Conservative 0; Mismatches 296; Indels 6; Gaps 2;

QY 124 GAGAAGTTGCGCCCTGAACCCCGGCTGCTGAGACCAAGGCGCTGCAAGCAGATCATC 183  
Db 282 GAGATCACCGCCCAACCAACCATGGAACCTGATGACGACGTGCGGAGCGAGGCGCAGTTT 341  
QY 184 CGCAGCTGACCCCGCCCTGACAGCCGAGAGAGAGCTGAAGCCTGTTCAACACC 243  
Db 342 CTCAACATGCTGCTCAAGCTCATCGCGGCCCAAGAACCATGAGATCGGCGCTTAACAC 401  
QY 244 GTGGCCACCTGTACTGCGTGACAGAAAGATGAGTCCGCGACCAACGAGGCGCTG 303  
Db 402 GGCTACTCCCTCTCGCCACCGCGCTGCGCATCCCGACGACGCGACCATCTTGCCATG 461  
QY 304 GACAGATGAGAGAGAGCAAGCAAGTGCCAGAGAAATCAAGCAGGCGCGAGCGCC 363  
Db 462 GACATCAACCGCGAGAACTACGAGCTGGGCTGCGTGCATCGAAGAGCGCGCTGAGCG 521  
QY 364 GACAGGCGCAAGTGAGCCAGAACTACCCCATGTCAGAACTGCAAGGCGCAGATGCTG 423  
Db 522 CACAAGATGACTTCGCGAGGG---CCCGGCGTGGCGGTGCTGAGCGCGCTGAGAG 578  
QY 424 CACGAGCCATCAGCCCCCGCACCTGAACGCTGGGTGAAGTGAATGAGAGAGGCC 483  
Db 579 GACGAGGCCAACAACGACGACCTTGACTTCTGTGAGACGCGACCAAGACACTAC 638  
QY 484 TTGAGCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCGCAGAC 543  
Db 639 CTCAACTACACGAGCGCCTCATGAAGCTGTGAAGCTGCGGCGCTCTCGGCTACGAC 698  
QY 544 CTGAACACGATGTTGAACACCGTGGCGGCCACGAGCGCGCATGAGATGCTGAAGAC 603  
Db 699 ---AACAGCTGTGAACGCGTGGTGTGCTCCCGCGAGCGCCCATGCGCAAGTAC 755  
QY 604 ACCATCAACGAGAGCGCGCGAGTGGACCGCGTGACCCCGTGACCGCGCGCCCATC 663  
Db 756 ATCGCTACTACCGGACTTCTGCTCGACCTCAACAGGCGCTTGCGCGNCGACGCGC 815  
QY 664 GCCCGCGCCAGATGCGCGGAGCGCGCGCGGACGACATCGCGCGC 708  
Db 816 GTGAGATCTGCCAGCTCCCGTGGCGGACGCGATCACCTCTGC 860

RESULT 4  
CL968487 2277 bp DNA linear GSS 21-SEP-2004  
LOCUS OslFCC040489 Oryza sativa Express Library Oryza sativa (Indica  
DEFINITION cultivar-group) genomic, genomic survey sequence.  
ACCESSION CL968487  
VERSION CL968487.1 GI:52391607  
KEYWORDS GSS.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
REFERENCE Ma, L., Wang, C., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G.K.S., Deng, X.W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis  
Unpublished (2004)  
TITLE JOURNAL  
AUTHORS COMMENT  
CONTACT: Chen Chen  
Department of Bioinformatic

Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.

FEATURES  
source

location/Qualifiers  
1..2277  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Express Library"  
/note="Oryza sativa exon trapped genomic sequences "

ORIGIN

Query Match 5.4%; Score 80; DB 9; Length 2277;  
Best Local Similarity 42.2%; Pred. No. 0.00015;  
Matches 510; Conservative 0; Mismatches 695; Indels 3; Gaps 1;

QY 5 GCGCCGCGCCAGCATCTCTGCGCGCGCAAGCTGACGCTGGAGCGCATCCGCTGC 64  
Db 197 GCACAAACGACGCGATCTCTACGCGCGCGCGCGCAACAGAGGCGCGCGGTGA 256  
QY 65 GCGCCGCGCGCAAGAGTGTACATGATGAAGACACCTGTGTGGCGACCGCGAGCTGG 124  
Db 257 TCGCGCGGAGAGGCTGTCTACTTCTTACAGACATCTGTACGCGGTTTCAACGCGCGGC 316  
QY 125 AGAAGTTGCGCTGAAACCCCGCGCTGTGAGACCAAGGCGCTGCAAGCATATCC 184  
Db 317 TCAGGCGCGACGAGGCGCGCGCATGCGAGGCTGAGTGTGTCTGACGCGCTCCGG 376  
QY 185 GCCAGCTGACCCCGCGCTGACAGCCGCGAGGAGCTGAAAGCCTGTTCAACCG 244  
Db 377 ACTCAAGTACACCTTACACCAACGAGACAGCGCGCGCTGTCTGCGCATGAGCACCGCGC 436  
QY 245 TGGCCACCTGTACTGCTGTCAGCAGAGAGATCGAGTCCGCGACCAAGAGCGCTGG 304  
Db 437 GGACCGGCGCGTGGAGCGTGGCGCGGAAATGGCGGACGCGCTTCTGCGCTCTCG 496  
QY 305 ACAAGATCGAGGAGAGCAACAAGTCCAGCAGAGATCCAGAGCGCGAGCGCGCG 364  
Db 497 ACAACGCGCTGACCGCGCGCGCATCTCTGTTGCGGAGCAGAGGAGTGGCGCGCGCG 556  
QY 365 ACAAGGCGCAAGGTGAGCCAGAACTACCTGTCAGAACTGCGAGGCGCGAGATGTGC 424  
Db 557 CCAAGTGGCGCGGCAAGTGCAGATCTGCGCGCGCGCGTGCACAAAGCTCATCGCGC 616  
QY 425 ACCAGGCGCATAGCGCGCGCACCTGAAGCGCTGTGAAGGTGATCGAGAGAGCGCT 484  
Db 617 GCGGCGCGCAAGACGCTGAGAGACCAAGCAGCAGCATGTGGGCAACGCGCTGGCGCT 676  
QY 485 TCAGCCCGGAGGTATCCCATGTTCAACGCGCTGAGCGAGGCGCGCACCCCGCAGACC 544  
Db 677 TCGTCAGGAGCTCATGTGAGAGGCTCAACCTCGGAGCGCTTCGCGATGGCGCGCC 736  
QY 545 ---TGAACACGATGTGAACACCGTGGCGCGCGCACAGCGCGCGCATGAGATGTGAAG 601  
Db 737 GCGCGCACCTGGCGATGTACGAGGTGTGCTTCGCGACATGTGTCCGACAGGATGC 796  
QY 602 ACAACATCAACGAGAGGCGCGCGAGTGGAGCGCGTGCACCCCGTGCACGCGCGCGCA 661  
Db 797 TCACCGCGCAGAGAGGCGCGCTTCTCTGACGCGCTGACGCTCTTCCATCTCCGCT 856  
QY 662 TCGCCCGCGCGCAGATGCGCGAGCGCGCGCGCGAGCATGCGCGCGCACCAAGCAGACC 721  
Db 857 CCGACAACAAGCAGAAAGCCCTTCTAAGACGACTCATGCGCGTGGAGCTTCTCCGCG 916  
QY 722 TGCAGAGAGCATGCGCTGATGACAGCAACCCCGCATCCCGTGGCGAGATCTACA 781  
Db 917 TCATGGCGGCGCTTCTTTCAGCACAGCGCGGGAACGCGCGCGCACCGAGAGACAG 976  
QY 782 AGCGGTGATCATCTCTGGCGCTGAACAAGATCGTGGAGATGTACAGCCCGTGAAGCATCC 841

Db 977 TCACCAACTGCGCGCGCTGGAGCTACCGTCCGCGCGAGCAGCCTGGAGCGCGCATAA 1036  
QY 842 TGCATCATCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 901  
Db 1037 TCTCCAAGGTCCAGCTCTGGCGAGCGCGCTGTGTATCAACGCGCGCGCGCGCGTGA 1096  
QY 902 CCCTGCGCGCGCGCGAGAGCAGCAGGAGTGAAGAACTGATGACCGACCGCTGTGG 961  
Db 1097 AACCGCTCCAGAACAGCGCGATGTCTACGTGCGCGCGCGCGCGCGCGCGCGCGTGA 1156  
QY 962 TGCAGAACGCAACCCCGACTGCAAGACCATCTCGCGCGCTTCCGCGCGCGCGCGCG 1021  
Db 1157 AGCGGCTGACATCCGGGACAGATCGTGTCTGCAATCCGCTGAAAGTGCAGCCATGC 1216  
QY 1022 TGCAGAGATGATGACCGCGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGTGC 1081  
Db 1217 TCGAGAAAGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 1276  
QY 1082 TGGCCGAGGCGCATGAGCCAGGCGCAACCGAGCGTGTATGATGAGAGAACTTCAAG 1141  
Db 1277 AGTTCTCGCGCAACACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTACC 1336  
QY 1142 GCGCCCGCGCGCATCTGTAAGTCTTCACTGCGCGCGCGCGCGCGCGCGCGCGCGT 1201  
Db 1337 CCGACGCGCGAGCATCAAGGCGTATCACTCCACGCGCGCAACCCCATGCGGAGCTCC 1396  
QY 1202 GCGCGCGC 1209  
Db 1397 GCTTCCGC 1404

RESULT 5  
CK211108/c 1126 bp mRNA linear EST 09-DEC-2003  
LOCUS  
DEFINITION FGAS022942 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
VERSION  
ACCESSION CK211108  
KEYWORDS  
SOURCE  
ORGANISM  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1126)

REFERENCE  
AUTHORS  
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
Penniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress In Wheat and Canola Crops  
Unpublished (2003)  
CONTACT: Wm L. Crosby  
COMMENT

TITLE  
JOURNAL  
COMMENT  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1c101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_est@cs.usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [12,750].  
Plate: L6B001 row: A column: 20.

FEATURES  
source

1..1126  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown

(50%) and leaf (50%) tissues from wheat cultivar Norstar after short exposure times to low temperature in the light and in the dark. 12 mRNA populations were combined before constructing the library. The first 6 populations: After 7 days of growth at 20C from wheat cultivar Norstar after short exposure times to low temperature in the light and in the dark. 12 mRNA populations were combined before constructing the library. The first 6 populations: After 7 days of growth at 20, wheat plants were transferred to 4C in the light. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. The last 6 populations: After 7 days of growth at 20C, wheat plants were transferred to 4C in the dark. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI. In addition, this library used a primer for second strand synthesis that annealed to an artificial sequence (RNA oligo) added before first strand synthesis. Therefore when sequences from EST generated from this library will be masked for vector and adaptor sequences, an additional masking step will have to be included to mask this RNA oligo that is common to all clones (sequence CGACTGAGCAGCAGGACTGACATGACTGAAGAGTGAAGA)."

ORIGIN

Query Match 5.4%; Score 79.8; DB 7; Length 1126;  
 Best Local Similarity 48.2%; Pred. No. 0.00015;  
 Matches 287; Conservative 0; Mismatches 302; Indels 6; Gaps 2;

OY 124 GAGAGTTCGCTGACCCCGCTGAGAGCAGGAGGCTGCAAGCATCATC 183  
 DB 763 GAGATCAACCGCAACCCATGAACTGATGACGACGTCGCGAGGCGCACTTC 704  
 OY 184 CGCAGCTGACCCCGCTGACAGCGGAGGAGCTGAAGAGCTGTTCAACC 243  
 DB 703 CTCAACATGCTGCTCAAGCTCATCGGCGCAAGAACATGAGATCGCGCTACACC 644  
 OY 244 GTGGCCACCTGTACTGCTGACGAGAGATGAGTCCGACCAAGAGGCGCTG 303  
 DB 643 GGCTACTCCCTCCGCGACCGGCTGCGCATCCCGACGACGCGACCATCTTGGCCATG 584  
 OY 304 GACAAGATCGAGAGGAGCAAGAACTGTCAGAGATCCAGAGCGCGCGCC 363  
 DB 583 GACATCAACCGGAGAACTAGAGTGGGGCTGCGTGATGAGAGCGCGCTGAG 524  
 OY 364 GACAAGGCGAGTGAAGCACTACCTGTCAGAACCTGACGGGCGAGATGTTG 423  
 DB 523 CACAAGATGACTTCGCGAGAGG---CCCGGCGCTGCGGTGTCGAGCGCTGTCGAG 467  
 OY 424 CACGAGGCGATCAGCCCGGACCTGAGAGCGCTGGTGAAGTGAATGAGAGAGGCC 483  
 DB 466 GACGAGGCGCAACGAGCACTTCACTTCTTGTGAGCGCGCAAGACCACTAC 407  
 OY 484 TTGAGCCCGGAGGTATCCCATGTTCAACCGCTGAGCGAGGCGCGACCCCGAGAC 543  
 DB 406 CTCAACTACGAGCGCTCATGAGTCTGTCAGAGCTCGGCGCTCTGCTGACAG 347  
 OY 544 CTGACACGATGTTGAACACCGTGGGCGGCGACAGCGCGCATGAGATGTCAGAGAC 603  
 DB 346 ---AACAAGCTCTGAGAGGCTCCGCTGTGCTCCCGCGAGCGCCCATGCGCAATAC 290  
 OY 604 ACCATCAACGAGAGGCGCGGAGTGGAGCGCGCTGACCCCGTGAACGCGCGCGCCATC 663  
 DB 289 ATCGCTCTTCAACGAGCTTCTGCTGACCTCAACAGGCGCTCGCGCGAGCCAGCGCG 230  
 OY 664 GCGCGCGCGAGATGCGCGAGCGCGCGCGAGCATCGCGCGCAACCAAGCA 718  
 DB 229 GTGAGATCTGCGAGCTCCCGTGGGCGAGCGGATCACTCTGCGCGCGCGCA 175

RESULT 6  
 CN128942  
 LOCUS  
 DEFINITION  
 RHOH1\_32\_C05\_g1\_A002 Acid- and alkaline-treated roots Sorghum bicolor cDNA clone RHOH1\_32\_C05\_A002 5', mRNA sequence.  
 CN128942  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Sorghum bicolor (sorghum)  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 826)  
 Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Lim,S., Eastman,A. and Pratt,L.H.  
 An EST database from Sorghum: acid- and alkaline-treated roots Unpublished (2003)  
 Other ESTs: RHOH1\_32\_C05\_b1\_A002  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sugs (CTTCTGCTTAAAGCTGCG).

FEATURES

source  
 1..826  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /cultivar="BTx623"  
 /db\_xref="taxon:4558"  
 /clone="RHOH1\_32\_C05\_A002"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /note="Organ: Root; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from 8-day-old roots harvested from BTx623 sorghum seedlings grown in hydroponic culture. HCl was added to a pH of 3.0 to some seedlings, KOH to a pH of 9.0 for others. Roots were harvested 3, 12 and 27 hr after addition and pooled for RNA preparation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Query Match 5.4%; Score 79.4; DB 7; Length 826;  
 Best Local Similarity 48.3%; Pred. No. 0.00018;  
 Matches 321; Conservative 0; Mismatches 331; Indels 13; Gaps 3;

OY 102 GGTGTGGGCGAGCGGAGCTGAGAAAGTTCCGCTGAAACCCGCTGTCGAGACAG 161  
 DB 160 GGTGAGACACAAGAGCCTGCTCAAGAGCGAGCACTCTACAGTACATCTGACACAG 219  
 OY 162 CGAGGCTGCAAGCATCATCGCGAGCTGACCCCGCTGAGAGCGCGAGAGGA 221  
 DB 220 CTTGTACCCCGGAGAGCGGAGAGCATGAGAGCTCCGCGAGATCACCGCGAGACCC 279  
 OY 222 GCTGAAGAGCTGTTCAACACCGTGGCGACCTGTACTGCGTGCAGAGAGATCGAGT 281  
 DB 280 ATGGAACCTGATGACGACTCCGCGAGCGAGGCGAGTTCCTCAACATGCTCAAGCT 339

```
QY 282 CCGCGACACCAAGAGGCCCCCTGGACAGATCGAGAGGACGAAACAAGTCCAGCAGAA 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 CATGGCGCCCAAGAGAACCATGAGCATCAACCGCGAAGACTACGAGCTGGCCTGCTG 399
QY 342 GATTCAGCAGGCGCGAGGCGCCGCGACAGAGGCGAAGGTGAGCCAGAACTACCCCATGTGCA 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 CATGAGAGAGGCGCGGCTGCGCCACAGATGACTTCCGCGAG-----GGCCCCGCGCT 453
QY 402 GAACCTGCAGGCGCAGATGTTGTCACAGGCGCATCAGCCCCCGCACCTGAAAGCCTGGGT 461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 454 CCGCGTCTCGACGACTCATCGCGGACGAGAGAACCAAGGTTGTTGACTTGTCTT 513
QY 462 GAAGGTGATCGAGAGAGAGGCTTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCTGAG 521
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 CGTGAACGCGGACAGAGCAACTACTCACTACCAAGACGAGCGCTGCTCAAGCTGGTGA 573
QY 522 CGAGGCGCGCACCCCCAGACCTGAAACAGATGTTGAACACCGTGGGCGGCCACAGGC 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 GCTGGGGGGCTCTCATCGGCTACGACAAACAGCTGTGAAACGGCTGTCTGCTCCCGA 633
QY 582 CGCCATGCAGATGCTGAAGAGACACCATCAACGAGAGGCGCGCAGTGGGACCGCGTGCA 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 CGACGCCCCCATGCGCAAGTACATCCGCTTCTA-----CCGCACTTCTGTCTGCTCT 687
QY 642 CCGCGTGACAGCGCGGCGCCCATCGCCCCCGGCGACATGCGCGAGCGCGCGGACGACAT 701
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 688 CAACAGGCGCTCGCGCGCGACGAGCGCTGAGATCTGCAAGCTTCCCGTGGCGACAG 747
QY 702 CGCGGACACCAACGACACCTGTCAGAGAGCAGATCGCCTGATGACAGCAACCCCGCAT 761
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 748 CGTCACTCTGCGCGCGCGT-CAAGTGAAGAAAAATGCTGCGCACCGAGCGCGCGC 806
QY 762 CCGCG 766
    ||| |||
Db 807 CGCG 811
```

```
RESULT 7          807 bp      mRNA      linear      EST 09-APR-2003
CB673622          OSJNE08D20.f OSJNE Oryza sativa (japonica cultivar-group) cDNA
LOCUS              clone OSJNE08D20 5', mRNA sequence.
DEFINITION
ACCESSION          CB673622
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 807)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tac gac cat g
Plate: 08      row: D      column: 20
Seq primer: gta aaa cga cgg cca gtcg.
FEATURES
source
1..807
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
```

```
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNE08D20"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNE"
/vector="pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; 24 hrs after immunoliation with Rice Blast (70-15)"
```

```
Query Match      5.3%; Score 78; DB 6; Length 807;
Best Local Similarity 44.4%; Pred. No. 0.00032;
Matches 312; Conservative 0; Mismatches 390; Indels 0; Gaps 0;

QY 204 GACAGCCGCGACGAGAGCTGAAGAGCTGTTCACACCCGTGGCCACCTGTACTGCGT 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 GCAGATGAAGATGGCGCGGAGACGTTCTCTTCACTCCGAGTCCGTGAACGAGGTCA 149
QY 264 GACAGAGAGATCGAGGTCCGCGACACCAAGAGAGGCCCTGGAACAAGATCGAGAGAGCA 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 CCGGACAAAGCTGTGCGACCAAGTGTCCGACGCGGTGCTGCAAGCGCTGCGCCCAAG 209
QY 324 GAACAGTGCACGACGAGAGATCCAGAGCGCGGAGGCCCGCGCAAGGGCAAGTGAGCCA 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 CCGGACAGCAAGGTGGCGGTGCGAGACGTGCAACCAAGACCAACATGATGTGTTCCG 269
QY 384 GAACCTACCCCATCTGTGACAGAACTTGACAGGGCCAGATGTTGCAACAGGCCATCAGCCCCG 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 CGAGATCACCAACCAAGGCCACCGTGCATCAAGAGATGTTCCGCGACACTGCGCGCG 329
QY 444 CACCCTGAAGCGCTGGGTGAAGTGTATCGAGAGAGAGGCGCTTCAAGCCCAAGTGATCCC 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 CATCGGCTGTGTGTCGACGACAGTCCGCTTCGACGCGCGCTGCAAGGTGCTGTCTCA 389
QY 504 CATGTTCAAGCGCCCTGAGCGAGAGGGCGCCACCCCGACGACTGAACACGATGTTGAACAC 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 CATCGAGCAGCAGTGCGCCGACATCGCGGAGGGGGTGACGCGCCACTTCAACCAAGCGCCC 449
QY 564 CGTGGCGGCGCACGAGCGCGCCATGACAGTGTGAAGAGACCATCAAGAGAGAGCGCGC 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 450 CGAGGAGATGCGCGCGCGGCGACCAAGGCCACATGTTCCGCTACGCCACCAAGACCCC 509
QY 624 CGAGTGGACCGCGTGCACCCCGTGCACGCGCGCGCCCATGCCCCCGGCGCAGATGCGCGCA 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 510 CGAGCTGATGCCCCCTCAGCCACGTCCTCGCCACCAAGCTTGCGCGCCCTCACCAGGT 569
QY 684 GCGCGCGGCGAGCGACATGCGCGGACACCAAGCAACCTGTCAGAGAGAGATGCGCTGGAT 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 570 CCGCAAGAGAGCGACCTGCGCGCTGAGCGCGCGCGCGCGCAAGCCAGGTCAACCTTGA 629
QY 744 GACCAAGCAACCCCGCATCCCGTGGGCGACATCTACAAGCGGTGATATCTTGCGCT 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 630 GTACCTCAACGACGCGCGCGCATGTGTTCCCGCTCCGCTCCACACCGTCTCATCTCCAC 689
QY 804 GAACAGATGCTGCGGATGTAACGCCCCGTGAGCATCTGGAACATCAAGAGAGGCGCCCA 863
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 690 CCAAGCAGCAGAGACCGGTCAACCAAGCAGAGATGCGCGCGCACTCAAGAGAGCAGTCA 749
QY 864 GAGGCCCTTCCGCGACTAGCTGAGCGGCTTCTTCAAGACCT 905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 750 CAAGCGGTATCCCCGACAGTACCTGACGAGAGACCAT 791
```

```
RESULT 8          821 bp      mRNA      linear      EST 09-APR-2003
CB662794          OSJNEd07C09.f OSJNEd Oryza sativa (japonica cultivar-group) cDNA
LOCUS              clone OSJNEd07C09 5', mRNA sequence.
DEFINITION
ACCESSION          CB662794
VERSION
KEYWORDS
SOURCE
Oryza sativa (japonica cultivar-group)
```

ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 821)  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
Unpublished (2003)  
JOURNAL  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aaa cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 07 row: C column: 09  
Seq primer: gta aaa cga cgg cca gtc.  
FEATURES  
source  
1..821  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSJNBd07C09"  
/issue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSJNBd"  
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"  
ORIGIN  
Query Match 5.3%; Score 78; DB 6; Length 821;  
Best Local Similarity 44.4%; Pred. No. 0.00032;  
Matches 312; Conservative 0; Mismatches 390; Indels 0; Gaps 0;  
QY 204 GCAGACCGGCGAGAGAGCTGAAGAGCTGTTCACACCGTGCCACCTGTACTGCGT 263  
DB 88 GCAGATAGAAATGGCGCGGAGACGTTCTTCACTCCGAGTCCGTGAACGAGGTC 147  
QY 264 GCAGAGAAGATCGAGTCCGCGACACCAAGAGCGCTTGACAAGATCGAGAGACA 323  
DB 148 CCGGACAAAGTGTGCGACGAGTGTGGAACGCGGTGCTCGACGCGCTCGCCAGGA 207  
QY 324 GAACAAGTCCAGCAGAAGATCCAGCAGCGCCGAGCGCCGACCAAGSGCAAGTGA 383  
DB 208 CCGGACAGCAAGTGGCGTGCAGACGTGCACCAAGACCAACATGTGATGTGTCG 267  
QY 384 GAACTACCCCATCTGTGAGAAGCTGACGAGGCGCAAGATGTGCACCAAGCCATCA 443  
DB 268 CGAGATCAACCAAGGCCACCGTCACTACGAGAATGTCCTCCGCAACCTGCGCG 327  
QY 444 CACCTGAACGCGCTGGGTGAAGTGAATGAGAGAGAGCGCTTCAAGCCCGAGTGA 503  
DB 328 CATCGCTTCTGTGTCGACGACGTGCGCTCGACGCGCAACGCTGCAAGTGTCTCA 387  
QY 504 CATGTTACCGCCCTGAGCGAGGCGGCCACCCCGCAGACCTGAACAAGATGTTGA 563  
DB 388 CATCGACGACGATCGCGCCGACATCGCGCAGGGGTGACGGCACTTCAACCAAG 447  
QY 564 CGTGGCGGCCACCAAGCGCCCATGACATGCTGAAGACACCATCAACGAGAGCGCC 623  
DB 448 CGAGGAGATCGCGCGCGGCGACCAAGGCGCAATGTTGCGCTACGCCACCGAGAC 507  
QY 624 CGAGTGGAGACCGCGCTGACACCCCGTGAACGCGCGCCCATTCGCCCGCAGATG 683  
DB 508 CGAGCTGATGCCCCCTCAACGACGTCCTGCGCACCAAGCTCGGCGCGCTCAACGA 567

QY 684 GCCCGCGGACGACATCGCCCGCACCAACGACCCCTGCAGAGCAGATCGCTGGAT 743  
DB 568 CCGCAAGAACGGCACTGCGCTGGCTCAGAGCCCGACGCAAGACCGAGTCACTGGA 627  
QY 744 GACCAAGACCCCGCCATCCCGGTGGCGGACATCTAACAGCGGTGATCATCTGG 803  
DB 628 GTACCTCAACGAGCGCGCGGCGCATGTGCCCGTCCGCTCCACACCTCATCTCA 687  
QY 804 GAACAAGATCGTGGGATGTACAGCCCGGTGACATCTGACATCAAGCAGAGGCC 863  
DB 688 CCAGCAGCAGCAGACCGCTCACCAAGCAGAGATCGCCGCGACTCAAGAGCAGT 747  
QY 864 GAGCCCTTCCGCACTACGTGACCGCTTCTTCAAGACCT 905  
DB 748 CAGCCGCTCATCCCGCAGCAAGTACCTCGACGAGAAGACCAT 789  
RESULT 9  
CB666192 824 bp mRNA linear EST 09-APR-2003  
LOCUS  
DEFINITION OSJNBd12P19.f OSJNBd Oryza sativa (japonica cultivar-group) cDNA  
clone OSJNBd12P19 5', mRNA sequence.  
ACCESSION  
CB666192 GI:29669917  
VERSION  
CB666192  
KEYWORDS  
SOURCE  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 824)  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
Unpublished (2003)  
JOURNAL  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aaa cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 12 row: P column: 19  
Seq primer: gta aaa cga cgg cca gtc.  
FEATURES  
source  
1..824  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSJNBd12P19"  
/issue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSJNBd"  
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"  
ORIGIN  
Query Match 5.3%; Score 78; DB 6; Length 824;  
Best Local Similarity 44.4%; Pred. No. 0.00032;  
Matches 312; Conservative 0; Mismatches 390; Indels 0; Gaps 0;  
QY 204 GCAGACCGGCGAGAGCTGAAGAGCTGTTCACACCGTGCCACCTGTACTGCGT 263  
DB 98 GCAGATAGAAATGGCGCGGAGACGTTCTTCACTCCGAGTCCGTGAACGAGGTC 157

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	264	GCACGAGAAAGATGAGGTC	CCGCGACACCAAGGAGCCCTTGACAAAGATCGAGGAGCA	323				
Db	158	CCGGGACAAAGCTGTGCGAC	CAAGGTGTGGACCGGTGCTCGACCGGTGCTCGCCAGGA	217				
Db	324	GAAACAAGTGCCAGAGAAAGAT	CCAGCAGGCGCCCGACAAAGGGCAAGGTGAGCCA	383				
Db	218	CCCCGACAGCAAGGTGGCGTG	CAGACGTGCACCAAGAACCAACATGATGATGTTCCG	277				
Db	384	GAACTACCCCATCGTGCAGAA	ACCTTCAGAGGGCCAGATGGTGACACGAGCCATCAGCCCCG	443				
Db	278	CGAGATCACCACCAAGGCCA	CCGTCGACTACGAGAAAGATCGTCCCGGACACCTGCGCGG	337				
Db	444	CACCTGAAACGCGCTGGGGT	GAAAGTGATCGAGGAGAAAGCCCTTCAGCCCCGAGGTGATCCC	503				
Db	338	CATCGGCTTCGTGTCGACGA	CGTGGGCTTCGACGCCGACCGCTTCAGAGGTGCTGTCGCA	397				
Db	504	CATGTTACCCGCTGAGCCG	AGGGCCACCCCGCAGGACCTTGAACGATGTTGAACAC	563				
Db	398	CATCGAGCAGCAGTCCGCC	CGACATCGCGAGGGGTGCACGGCCACTTCACCAAGCGCCC	457				
Db	564	CGTGGGGCGCCACCAAGG	CGCCGATGCTGAAGACACCATCAACGAGAGCGCCG	623				
Db	458	CGAGAGATCGGCGCGCGCG	ACCAAGGACACATGTTGGCTACGCCACCGAGAGACCCC	517				
Db	624	CGAGTGGGACCGCGTGCA	CCCCGTCGACGCGGCCCCATCGCCCCCGCCAGATGCGCGA	683				
Db	518	CGAGCTGATGCCCTCAGCC	ACGTCCTCGCCACCAAGCTGGCGCGCCTCACCGAGGT	577				
Db	684	GCCCCCGCGCAGCAGCAT	CGCCGCGACCAAGACCCCTGCAGAGCAGATCGCCTGAT	743				
Db	578	CCGCAAGAACGGCACTGC	CGCTGCTCAGGCCCGACGGCAAGACCAGTACCGTTGA	637				
Db	744	GACCAAGCAACCCCCCAT	CCCCGTCGAGCATCAAGCGGTGATCATCTGGGCT	803				
Db	638	GTAACCTCAACGACGCG	CGCCGATGTTCCCGTCGCTCCACACCGTCTCATCTCCAC	697				
Db	804	GAAACAAGATCGTGGGAT	GTACAGCCCCGTGAGCATCTTGACATCAAGCAGGCGCCAA	863				
Db	698	CCAGCAGCAGCAGACCG	CTCACCAAGCAGAGATGCGCCGACCTCAAGAGACAGTCAT	757				
Db	864	GGAGCCCTTCCGCGACT	ACGTGACCGCTTCTTCAAGACCT	905				
Db	758	CAGCCCGTCAATCCCGA	CAAGTACCTCGACGAGAGACCAT	799				
RESULT 10								
LOCUS	CB645622							
DEFINITION	OSJNB07H03.f OSJNB	Oryza sativa (japonica cultivar-group)	CDNA					
ACCESSION	CB645622							
VERSION	CB645622.1	GI:29640613						
KEYWORDS	EST.							
SOURCE	Oryza sativa (japonica	cultivar-group)						
ORGANISM	Oryza sativa (japonica	cultivar-group)						
REFERENCE	1 (bases 1 to 851)							
AUTHORS	Jantasuriyarat,C., Lu,G.,	Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,						
TITLE	Kudra,D., Dean,R., Soderlund,C.,	Wing,R. and Wang,G.						
JOURNAL	Large-scale identification of ESTs	involved in the interaction						
COMMENT	Unpublished (2003)							
	Contact: Rod Wing							
	Arizona Genomics Institute							
	University of Arizona							
	Biological Sciences West, 448A,	P.O. Box 210088, Tucson, AZ						
	85721-0088, USA							
	Tel: 520 626 3967							
	Fax: 520 621 9288							
	Email: http://genome.arizona.edu							
	PCR Primers							

```

FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 07 row: H column: 03
Seq primer: gta aaa cga cgg cca gtc
Location/Qualifiers
1. 851

```

```

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB07H03"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNB"
/note="Vector: pBluescript II KS+; Site_1: EcoRI; Site_2:
XhoI; 24 hrs after inoculation with Rice Blast (Che
86061)"

```

Query Match	5.3%;	Score 78;	DB 6;	Length 851;
Best Local Similarity	44.4%;	Pred. No. 0.00032;		
Matches 312; Conservative	0;	Mismatches 390;	Indels 0;	Gaps 0;

QY	204	GCAGACCCGGCAGCGAGGAGCTGAAAGAGCCTGTTCAACACCCGTGGCCACCCCTGTACTGCCG	263
Db	105	GCAGATAGAAATGGCCGGCGAGACGTTCCCTCTTCACTCCGAGTCCGTGAAACGAGGGTCA	164
QY	264	GCACGAGAAGATCGAGGTCCCGGACACCAAGAGAGGCCCTGGACAAGATCGAGAGAGACA	323
Db	165	CCCCGACCAAGCTGTGCCGACCAAGGTTCGGACCGGGTGTCTGCACGGCTGCTGCCACAGA	224
QY	324	GAACAAGTGCACGAGAAGATCCAGCAGGCGGAGGCCCGCCGACAAAGGGCAAGGTGAGCCA	383
Db	225	CCCCGACAGCAAGGTGGCGTGGAGACGTGCACCAAGACCAACATGTTGTTGTTCCG	284
QY	384	GAACACTACCCCATCTGTGCAGAACCTGCAGGGGCGAGTGTGCACCAAGCCATCAGCCCCCG	443
Db	285	CGAGATCACCAACCAAGGCCACCGCTGACTAGAGAAGATCGTCCGGCACACTGCGCGG	344
QY	444	CACCCTGAAAGCCTGGGTGAAGTGTATCGAGAGAGAAGGCTTCAGCCCCGAGGTGATCCC	503
Db	345	CATCGGCTTGCTGTCCGACGACGTGGGCTCGACGCCGACCGCTGCAGAGTCTCGTCAA	404
QY	504	CATGTTCAACGCCCCCTGAGCGAGGGGGCCACCCCCCAGACCTGAAACAGATGTTGAACAC	563
Db	405	CATCGAGCAGCAGTCCGCCGACATCGCGAGGGGGTGCACGGCCACTTCACCAAGCGCCC	464
QY	564	CGTGGGCGGCCACCAAGCGCGGCATGCAGATGCTGAAGACAACATCAACGAGAGGCGCG	623
Db	465	CGAGGAGATGGGCGCGGCGCACAGGCGCATGTTCCGCTTACGCCACCGACGAGACCCC	524
QY	624	CGAGTGGAGCCGCGTGCAACCCCGTGCAAGCGCGGCCCATCGCCCCCGGCAGATGCGGA	683
Db	525	CGAGCTGATGCCCTCAGCCACGTCTCTGCCACCAAGCTCGGCGCCCCGCTCACCGAGGT	584
QY	684	GCCCCGCGGACGACATCGCGCGGCACCAACAGCACCCCTGCAGAGAGCAGATCGCCTGAT	743
Db	585	CCGCAAGAACGGCACCTCGCGCTGGCTCAGGCCCGCAGCGCAAGACCAGGTCAACGGTTGA	644
QY	744	GACCAGCAACCCCCCATCCCGCTGGGCGACATCTACAAGCGGTGATCATCTGGGCTT	803
Db	645	GTACCTCAACGACGCGGCGGCATGTCTCCGCTCCGCTCACACCGTCTCATCTCCAC	704
QY	804	GAACAAGATCGTGCAGATGTACAGCCCCGTGAGCATCTTGACATCAAGCAGGCCCCCAA	863
Db	705	CCAGCACGACGAGACCGTCAACCAACGACGAGATCGCGCGGACCTCAAGAGACAGTCA	764
QY	864	GGAGCCCTTCCGCGACTACGTGAGACCGCTTCTTCAAGACCT	905
Db	765	CAAGCCGCTATCCCGCAGCAGTACTCTGCACGAGAAGACCT	806

```
RESULT 11
CB673355          807 bp      mRNA      linear      EST 09-APR-2003
LOCUS             OSJNE07N10.f OSJNE Oryza sativa (japonica cultivar-group) cDNA
DEFINITION        clone OSJNE07N10 5', mRNA sequence.
ACCESSION         CB673355
VERSION           CB673355.1  GI:29677080
KEYWORDS          EST.
SOURCE            Oryza sativa (japonica cultivar-group)
ORGANISM          Oryza sativa (japonica cultivar-group)
                  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                  Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE         Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
                  Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
                  Large-scale identification of ESTs involved in the interaction
                  between rice and Magnaporthe grisea
                  Unpublished (2003)
JOURNAL           Contact: Rod Wing
                  Arizona Genomics Institute
                  University of Arizona
                  Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
                  85721-0088, USA
                  Tel: 520 626 3967
                  Fax: 520 621 9288
                  Email: http://genome.arizona.edu
PCR PRIMERS      FORWARD: gta aaa cga cgg cca gtc
                  BACKWARD: gga aac agc tat gac cat g
                  Plate: 07  row: N  column: 10
                  Seq primer: gta aaa cga cgg cca gtc.
FEATURES
  source          1..807
                  /organism="Oryza sativa (japonica cultivar-group)"
                  /mol_type="mRNA"
                  /cultivar="Nipponbare"
                  /db_xref="taxon:39947"
                  /clone="OSJNE07N10"
                  /tissue_type="Leaf"
                  /dev_stage="3 week"
                  /lab_host="DH10B"
                  /clone_lib="OSJNE"
                  /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
                  XhoI; 24 hrs after inoculation with Rice Blast (70-15)"
ORIGIN
Query Match      5.2%; Score 76.6; DB 6; Length 807;
Best local Similarity 44.3%; Pred. No. 0.00058;
Matches 310; Conservative 0; Mismatches 389; Indels 0; Gaps 0;
QY 204 GCAGACCGGCGAGGAGCTGAAGAGCTGTCAACACCGTGCCACCTGTACTGCGT 263
    |||||
DB 109 GCAGATAGAATAGCGCGGAGACGTTCTCTCACTCCGAGTCCGTGAACGAGGGTCA 168
QY 264 GCACGAGAAGATCGAGGTCCGCGACACCAAGAGGCGCTTGACAAGATCGAGAGGACA 323
    |||||
DB 169 CCCGGAACAAGTGTGCGACCAAGGTGTGAGACGCGGTGCTGACGCGCTCGCCCGAGA 228
QY 324 GAACAAAGTCCAGCAGAAAGATCCAGCAGGCGCGCGCCGACCAAGGGCAAGGTAGCCA 383
    |||||
DB 229 CCCCAGACAGCAAGGTGCGTGCAGAGCTGCACCAAGACCAACATGATGATGTTCCG 288
QY 384 GAATAACCCCATCGTCAGAACCTGCAGGGCCAGATGTCACACGAGCCATCAGCCCCCG 443
    |||||
DB 289 CGAGATCACCAACCAAGGCCACCGTCGACTACGAGAAGATCGTCCGCGACACCTGCCCGG 348
QY 444 CACCTGAACGCGCTGGGTGAAGGTGATCGAGAGAAAGGCTTCAGCCCCGAGGTGATCCC 503
    |||||
DB 349 CATCGGCTTCGTGTCCGACGACGTCGGGCTTCGACGCCGACCGCTGCAAGGTGCTCGTCAA 408
QY 504 CATGTTACCGCCCTGAGCGAGGGCGGCCACCCCGACGACCTGAACACGATGTTGAACAC 563
    |||||
```

```
DB 409 CATGAGCAGCAGTCCGCCGACATCGCGGAGGGGTGCACGCGCACTTCACCAAGCGCCC 468
QY 564 CGTGGCGGCCACCAAGCGCCGCAATGCAGATGCTGAAGACACCATCAAGAGAGCGCCG 623
    |||||
DB 469 CGAGGAGATCGGCGCGCGCGACCAAGGGCCACATATGTCGCTACGCCACCGAGACCCCC 528
QY 624 CGAGTGGGACCGGCTGCACCCCGTGCACGCGCGGCCCATCGCCCCCGGCAGATGCGCGA 683
    |||||
DB 529 CGAGCTGATGCCCCCTCAGCCACGCTCTCGCCACCAAGCTCGGCGCCGCTCACCAGAGT 588
QY 684 GCGCCCGCGGACGACATCGCGCGGACCAACGACCCCTGCAGAGACAGATCGCTTGAT 743
    |||||
DB 589 CCGCAAGAACCGGACCTGCGCTGCTCAGGCGCCGACGCAAGACCAAGTACCGTTGA 648
QY 744 GACCAACAACCCCCCATCCCGTGGGGCGACATCTACAAGCGGTGATCATCTGCGCCT 803
    |||||
DB 649 GTACCTCAACGACGCGCGCGCGCATGTCCTCCGTCGCTCCACACCGTCTCATCTCCAC 708
QY 804 GAACAAGATCGTGGGATGTACAGCCCCGTGAGCATCTGACATCAAGCAGGCCCCCAA 863
    |||||
DB 709 CCAGCAGCAGCAGACCGCTCACCAACGACGAGATCGCCGCGACCTCAAGAGCAGCTCAT 768
QY 864 GGAGCCCTTCGCGCACTACGTGAGCCGCTTCTTCAAGAC 902
    |||||
DB 769 CAAGCCGCTCATCCCCGACAAGTACTCTGACGAGAAAGAC 807

RESULT 12
CK163513          1138 bp      mRNA      linear      EST 05-DEC-2003
LOCUS             FGAS016142 Triticum aestivum FGAS: Library 4 Gate 8 Triticum
DEFINITION        aestivum cDNA, mRNA sequence.
ACCESSION         CK163513
VERSION           CK163513.1  GI:38993817
KEYWORDS          EST.
SOURCE            Triticum aestivum (bread wheat)
ORGANISM          Triticum aestivum
                  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                  Poideae; Triticeae; Triticum.
REFERENCE         Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
                  Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
                  Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
                  Penniket,C., Roach,J.L. and Sarhan,F.
                  Functional Genomics of Abiotic Stresses In Wheat and Canola Crops
                  Unpublished (2003)
JOURNAL           Contact: Wm L Crosby
                  Bioinformatics
                  University of Saskatchewan, Department of Computer Science
                  1C101 Engineering Building, 57 Campus Drive, Saskatoon,
                  Saskatchewan, S7N 5A9, Canada
                  Tel: 306 966 1769
                  Fax: 306 966 2033
                  Email: fgas_estes@cs.usask.ca
                  This sequence is the direct result of the Base calling software
                  Phred (default parameters). It is the raw base calls. To aid in the
                  identification of the high quality insert the software Lucy
                  (default parameters) has been run on this sequence. Lucy identified
                  the region [8,827].
                  Plate: L4B008  row: F  column: 04.
FEATURES
  source          1..1138
                  /organism="Triticum aestivum"
                  /mol_type="mRNA"
                  /db_xref="taxon:4565"
                  /clone_lib="Triticum aestivum FGAS: Library 4 Gate 8"
                  /note="Organ: Crown and leaf; Vector: PCMV.SPORT6;
                  Conditions for growth: Seeds were germinated in a
                  water-saturated mix (50% black earth and 50% Promix) in a
                  growth chamber for 7 days under an irradiance of 200 mmol
                  m-2 sec-1. The temperature was maintained at 20 degrees C
                  with a 15-hr photoperiod under a relative humidity of 70%.
```

After this period watering of plants was stopped. Four time points were sampled during a two week period; the first after wilting was observed and the last, two weeks later, consisted of live crown and leaf tissue (leaf tissue that was yellow was not included in sampled material). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

ORIGIN

Query Match 5.2%; Score 76.6; DB 7; Length 1138;  
Best Local Similarity 49.0%; Pred. No. 0.00059;  
Matches 291; Conservative 0; Mismatches 295; Indels 8; Gaps 3;

OY 124 GAGAGTTGCGCCCTGAACCCCGGCTGTGAGACGAGGGCTGCAAGCATCATC 183  
DB 232 GAGATCACCGCCCAACCCCATGAACTGTAGACGCTCGCGGACGAGGGCCAGTTT 291  
OY 184 CGCCAGCTGCACCCCGCCCTGACAGCGGACGAGAGCTGAAGCCTGTTCAACAC 243  
DB 292 CTCAACATGCTGTCAAGCTCATCGCGCCCAAGAACCATGAGATCGCGCTTACACC 351  
OY 244 GTGGCCACCTGTACTGCTGACAGAAAGATCGAGTCCGCGACCAAGAGGCGCTG 303  
DB 352 GGCTACTCCCTGCTGCGCCACCGCGCTGCCATCCCGACGACGACCATCTTGCCATG 411  
OY 304 GACAGATCGAGGAGGACGAGAACAGTGCCAGCAGAAAGATCCAGCAGCGCGCGCC 363  
DB 412 GACATCAACCGGAGAACTACGAGCTGGGGCTGCCGTGATCGAAGAGCGCGCTGGCG 471  
OY 364 GACAGGGGCAAGGTGACGACGAACTACCCCATCGTCAAGAACCTGACGAGGCGCAAGTGTG 423  
DB 472 CACAGATGCACTTCGCGAG--GCGCGCGCGCTGCCGTGATCGAGCGCGCTGTGGAG 528  
OY 424 CACCAAGCCATCAGCCCGCCGACCTGAACGCTGGGTGAAGTGAAGTGAAGAGGCC 483  
DB 529 GACGAGGCCAACCAAGGACCTTGCACTTCTGTTGAGCGCGCAAGAACACTAC 588  
OY 484 TTGACCCCGAGGTGATCCCATGTTTCAACCGCTGAGCGAGGGCGCCACCCCGAGAG 543  
DB 589 CTCACTACCAAGGCGCTCATGAACTCTGCAAGTCCGCGGCTCTCGGCTACGAC 648  
OY 544 CTGAACACGATGTTGAACACCGTGGCGGCGCACGAGCGCCCATGAGATGTTGAAGAC 603  
DB 649 --AACACCCCTTGGAACGGCTCCGCTGTGCTCCCGCGGACGCGCCCATGCGCAAGTAC 705  
OY 604 ACCATCAACGAGAGCGCGCGAGTGGAGCGC--GTGACCCCGCTGACAGCGCGGCCCA 661  
DB 706 ATCGCTACTACCGGCACTTCTCTCGAGCTCAACAGGCGCTCGCGCGNCGACGAGCGG 765  
OY 662 TCGCCCCCGGCAAGTGGCGGAGCGCGCGGACGACATCGCGGACCAACCA 715  
DB 766 TCGAGATCTGCGAGCTCCCGCTCGGCGAGCGGACATCACTCTGCGCGCGCGCCA 819

RESULT 13  
LOCUS CB681973 812 bp mRNA linear EST 09-APR-2003  
DEFINITION OSJNEf08M09.f OSJNEf Oryza sativa (japonica cultivar-group) cDNA  
ACCESSION CB681973  
VERSION CB681973  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

CB681973 812 bp mRNA linear EST 09-APR-2003  
OSJNEf08M09.f OSJNEf Oryza sativa (japonica cultivar-group) cDNA  
clone OSJNEf08M09 5', mRNA sequence.  
CB681973  
GI:29685698  
EST.  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 812)  
Jantaauriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea

JOURNAL

Unpublished (2003)  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: <http://genome.arizona.edu>

COMMENT

PCR Primers  
FORWARD: gta aaa cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 08 row: M column: 09  
Seq primer: gta aaa cga cgg cca gtc.  
Location/Qualifiers

FEATURES

source

1. 812  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSJNEf08M09"  
/issue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_1ib="OSJNEf"  
/note="Vector: Bluescript II KS +; Site\_1: EcoRI; Site\_2:  
XhoI; Uninfected Control"

ORIGIN

Query Match 5.2%; Score 76.4; DB 6; Length 812;  
Best Local Similarity 44.3%; Pred. No. 0.00064;  
Matches 311; Conservative 0; Mismatches 391; Indels 0; Gaps 0;

OY 204 GCAGACCGGCGAGGAGAGCTGAAGAGCTGTTCACACCGGTGGCCACCTGTACTGCGT 263  
DB 105 GCAGATGAATGGCGCGGAGACGTTCTCTTCACTCCAGTCCGTGAACGAGGTCA 164  
OY 264 GCAGGAGAGATCGAGGTCCCGGACACCAAGAGGCGCCCTGACAAAGATCGAGAGACA 323  
DB 165 CCGGACAAAGCTGTGCGACCAAGTGTCCGACGCGGTGCTCGACGCGTCCGCCAAGA 224  
OY 324 GAACAAAGTCCAGCAGAAAGATCCAGAGCGCGGAGCGCGGCAAGGGCAAGGTGAGCCA 383  
DB 225 CCGGACAGCAAGGTGGCGGTGCGAGACGTGCAACCAAGACCAATGATGTGTTCCG 284  
OY 384 GAACCTACCCATCGTGTGACAACTGACGGGCGAAGATGTGACACGAGCCATCAGCCCGC 443  
DB 285 CGAGATCACCAACCAAGGCGCACCTGTGACTACGAGAGATCGTCCGACACCTGCGCGG 344  
OY 444 CACCCTGAACGCGCTGGGTGAAGTGTGATCGAGAGAGAGGCGCTTCAAGCCCGAGTGTCC 503  
DB 345 CATGCGCTTGTGTCCGACGACGTGCGCTTGAACGCGGCTGCAAGGTGCTGTCAA 404  
OY 504 CATGTTACCGCCCTGAGCGAGGGCGCCACCCCGGAGCCTGAACACGATGTTGAACAC 563  
DB 405 CATGAGCAGCAGTCCCGGACATCGCGAGGGGGTGACGGGCCACTTACCAAGCGGCC 464  
OY 564 CGTGGCGGCGCACGAGCGCGCATGACAGTCTGAAGAGACCATCAACGAGGAGCGCGC 623  
DB 465 CGAGGAGATGCGCGCGCGGACCAAGGCGCAATGTTGCGCTTACGCCACGAGAGACCCC 524  
OY 624 CGAGTGGAGCGCGGTGACCCCGTGCACGCGCGGCGCCCATCGCCCCCGGCGCATGCGCGA 683  
DB 525 CGAGCTGATGCGCTTACGACCAAGTCTCGGCAACCAAGCTGGCGCGCGCTTCAACGAGGT 584  
OY 684 GCGCGCGGCGAGCATCGCGCGGACCAACGACGACGACGAGGAGAGAGAGAGAGTGGCTGAT 743  
DB 585 CCGCAAGAACGGCACCTGCGCTGCTGCTAGGCGCGGACGCGCAAGACCAAGTCAACGTTGA 644  
OY 744 GACCAAGACCCCGCATCCCGTGGGCGCATCTACAAGCGGTGATCATCTGGGCGCT 803  
DB 645 GTACCTCAACGACGCGCGGCGCATGTCCCGTCCGCTTCAACACCGTCTCATCTCCAC 704

QY	804	GAAACAAGATCGTGGCGATGTACAGCCCCGTGAGCATCTTGACATCAAGCAGGCCCCAA	863
Db	705	CCAGCAGCAGAGACCGCTCACCAACGACGAGATGCGCCGCACTCAAGAGCAGTCAT	764
QY	864	GGAGCCCTTCCGCGACTACGTGAGCCGCTTCTTCAAGACCT	905
Db	765	CAAGCCGCTATTCCCGACAAGTACTCTGACGAGAGAACCAT	806
RESULT 14			
CB684834			
LOCUS			
DEFINITION			
OSJNEf14L24.f OSJNEf Oryza sativa (japonica cultivar-group) CDNA			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
EST.			
Oryza sativa (japonica cultivar-group)			
Oryza sativa (japonica cultivar-group)			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
Ehrhartoideae; Oryzeae; Oryza.			
1 (bases 1 to 832)			
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,			
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.			
Large-scale identification of ESTs involved in the interaction			
between rice and Magnaporthe grisea			
Unpublished (2003)			
Contact: Rod Wing			
Arizona Genomics Institute			
University of Arizona			
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ			
85721-0088, USA			
Tel: 520 626 3967			
Fax: 520 621 9288			
Email: http://genome.arizona.edu			
PCR Primers			
FORWARD: gta aac cga cgg cca gtcg			
BACKWARD: gga aac agc tat gac cat g			
Plate: 14 row: L column: 24			
Seq primer: gta aac cga cgg cca gtcg.			
Location/Qualifiers			
1..832			
/organism="Oryza sativa (japonica cultivar-group)"			
/mol_type="mRNA"			
/cultivar="Nipponbare"			
/db_xref="taxon:39947"			
/clone="OSJNEf14L24"			
/issue_type="Leaf"			
/dev_stage="3 week"			
/lab_host="DH10B"			
/clone_lib="OSJNEf"			
/note="Vector: Bluescript II KS +; Site_1: EcoRI; Site_2:			
XhoI; Uninfected Control"			
ORIGIN			
Query Match			
Best Local Similarity 5.2%; Score 76.4; DB 6; Length 832;			
Matches 311; Conservative 0; Mismatches 391; Indels 0; Gaps 0;			
QY	204	GCAGACCCGCGAGCGAGAGCTGTAAGACCTGTTCACACCGTGCCACCTGTAAGCTGCT	263
Db	97	GCAGATAGAAATGGCGCGGAGACGTTCTTTCACCTCCGAGTCCGTGAACGAGGTCA	156
QY	264	GCAAGAGAAATCGAGGTCCCGCACACCAAGAGAGCCCTTGACAAGATCGAGAGAGCA	323
Db	157	CCCGGACAAAGCTGTGCGACCAAGGTGTCGACGGGTGCTCGACGCGTCTCGCCACAGA	216
QY	324	GAAACAAGTCCGACGAGAAATCGACGAGGCGCGCGCGCGACCAAGGGCAAGGTGAGCCA	383
Db	217	CCCCGACAGCAAGGTGCGGTGCGAGACGTGACCAAGACCAACATGTGTATGTGTTGG	276
QY	384	GAACTACCCCATCTGTCAGAACTTCGAGGCGCGAGATGTGTCACCAAGGCATCAGCCCCCG	443

Db	277	CGAGATCACCACCAAGGCCACCGCTCGACTACGAGAAAGATCGTCCGCGACACCTGCCGGG	336
QY	444	CACCTGAAAGCCTGGGTGAAGTGATCGAGGAGAAAGCCTTCAAGCCCCGAGGTGATCCC	503
Db	337	CATCGGCTTCGTGTCCGACGACGTCGGCCTCGACGCCGACCGCTGCAAGGTGCTCTCA	396
QY	504	CATGTTACCGCCCTGAGCGAGGGCGCCACCCCCCAAGACCTTGAACACGATGTTGAACAC	563
Db	397	CATCGAGCAGCAGTCGCCCCGACATCGCGCAGGGGGGTGCACGGGCCATTCAACCAAGCCCC	456
QY	564	CGTGGGGCGCCACCAAGCGCGCATGTCAGATGCTGAAAGACACCATCAACGAGAGCGCCG	623
Db	457	CGAGAGATCGGCGCGCCGCGACCAAGGCCACATGTTGCGCTACGCCACCGACGAGACCCC	516
QY	624	CGAGTGGGACCGCGCTGCACCCCGCTGCACGCGCGCCCCCATCGCCCCCGGCGCATGCGCGA	683
Db	517	CGAGCTGATGCCCCCTCAGCCACGTCCTCGCCACCAAGCTCGGCGCCCCGCTCACCGAGGT	576
QY	684	GCCCCGGGCGAGCGACATCGCGCGGCACCAACGACCCCTGCGAGAGCAGATCGCCTGGAT	743
Db	577	CCGCAAGAACGGCACCTGCGCCTGGCTCAGGCGCCGACGCGCAAGCACCAGGTCACCGTTGA	636
QY	744	GACCAAGCAACCCCCCATCCCGCTGGGCGACATCTACAAAGCGGTGATCATCTGGGCT	803
Db	637	GTACTTCAACGACGCGCGGCGCCATGCTCCCGCTCGCGTCCACACCGTCTCATCTCCAC	696
QY	804	GAAACAAGATCGTGCAGATGTACAGCCCCGTGAGCATCTTGACATCAAGCAGGCGCCCCA	863
Db	697	CCAGCACAAGCAGACCGCTCACCAACGACGAGATCGCGCCGACTTCAAGAGCAGTCAAT	756
QY	864	GGAGCCCTTCCGCGACTACGTGAGACCGCTTCTTCAAGACCT	905
Db	757	CAGGCGGTCAATCCCCGACAAAGTACCTCGACGAGAAAGACCAT	798

RESULT	15
LOCUS	BF259495
DEFINITION	759 bp mRNA linear EST 22-OCT-2001 HVSMEF0019H19f Hordeum vulgare seedling root EST library HVCDNA00077 (Etiolated and unstressed) Hordeum vulgare subsp. vulgare cDNA clone HVSMEF0019H19f, mRNA sequence.
ACCESSION	BF259495
VERSION	BF259495.2 GI:13120022
KEYWORDS	EST.
SOURCE	Hordeum vulgare subsp. vulgare Hordeum vulgare subsp. vulgare
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticaceae; Hordeum. 1 (bases 1 to 759)
REFERENCE	Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton,R.D., Oates,R. and Main,D. Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library Unpublished (2001) On Nov 16, 2000 this sequence version replaced gi:1188608.
AUTHORS	Contact: Wing RA
TITLE	Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu
JOURNAL	Total hg bases = 388 Seq primer: AATTAACTTCCTAAGGG
COMMENT	High quality sequence stop: 758. Location/Qualifiers 1..759 /organism="Hordeum vulgare subsp. vulgare" /mol_type="mRNA" /cultivar="Morex"
FEATURES	
source	

/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="HVSMBF0019H19F"  
/issue\_type="Seedling root"  
/lab\_host="TJc121"  
/clone\_1b="Hordeum vulgare seedling root EST library  
HVCdNA007 (Etisolated and unstressed)"  
/note="vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
Seeds were surface sterilized then germinated under axenic  
conditions in the dark at room temperature on filter paper  
with water, nystatin and cefotaxime in covered  
crystallization dishes. Five-day old seedling roots were  
then harvested, total RNA was prepared, poly(A) RNA was  
purified, one primary unamplified cDNA library was made,  
and 1 million pfu were in vivo excised to give Bluescript  
SK(-) cDNA phagemids. These steps were performed in the TJ  
Close laboratory at the University of California,  
Riverside (Choi, Close, Fenton). Phagemids were plated and  
picked at the Clemson University Genomics Institute (CUGI)  
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA  
preparations, DNA sequencing and sequence analysis were  
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,  
Oates, Rambo, Main). The sequence has been trimmed to  
remove vector sequence and contains a minimum of 100 bases  
of phred value 20 or above. For more details on library  
preparation and sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 5.2%; Score 76.2; DB 2; Length 759;  
Best Local Similarity 47.7%; Pred. No. 0.00069;  
Matches 284; Conservative 0; Mismatches 305; Indels 6; Gaps 2;

QY 124 GAGAAGTTGGCCCTGAAACCCCGCTGTGAGACCGAGGGCTGCAAGATCATC 183  
Db 13 GAGATCACCGCCCAACCAACCATGAACTGATGACGACGTCGGCGGACGAGGCCAGTTTC 72  
QY 184 CGCCAGTGCACCCCGCCCTGCAGACCGGACGAGGAGCTGAAGAGCTTTCAACACC 243  
Db 73 CTCAACATGCTGCTCAAGCTCATCGGCGCAAGAACCATGAGATCGGCGTCTACACG 132  
QY 244 GTGGCCACCCTGTACTGCGTGCACGAAAGATCGAGTCCGCGACACCAAGAGGCCCTG 303  
Db 133 GGCTACTCCCTGCTGCGCAACCGCGCTGCCATCCCGACGACGACCATCTTGGCCATG 192  
QY 304 GACAAGATCGAGAGGAGCAAGAACAGTCCAGAGATCCAGACGCGCGAGGCCGCC 363  
Db 193 GACATCAACCGGAGAACTACGAGCTGGGGCTGCCGTGCATCGAAGAGCCGCGTGGCG 252  
QY 364 GACAAGGGCAAGTGAAGCAGAACTACCCCATCGTGAGAACTGCAAGGGCCAGATGGTG 423  
Db 253 CACAAGATCGACTTCGGGAGGGCCCGCGCTC---CCCGTCTGAGCGCCTCTCTGAG 309  
QY 424 CACCAAGGCATCAGCCCGCACCCCTGAAGCCTGGGTGAAGTGAAGGAGGCC 483  
Db 310 GACGAGGCCAACACGCGCACTTGACATTGCTTGTGACGCGCAAGACAACTAC 369  
QY 484 TTCAAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCCACCCCAAGAC 543  
Db 370 CTCAACTACCAAGAGCGCTCATGAAGCTGTCAAGTGGCGGCTCTCGGCTACGAC 429  
QY 544 CTGAACAGATGTTGAACACCGTGGCGGCCACCAAGCCGCATGACAGTCTGAAGGAC 603  
Db 430 ---AACACCCTCTGAAGGCTCCGTGTGCTCCCGCGACGCCCCCATGCGCAAGTAC 486  
QY 604 ACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGACACCCCGTGACGCGCGGCCCATC 663  
Db 487 ATCCGCTACTACCGGCACTTCGTCCTGACCTCAACAGGCCCTCGCGCGCGACCGCGC 546

QY 664 GCCCCCGCCAGATGCGCGAGACCCCGCGCGAGGACATCGCCGCGACCAACGACCA 718  
Db 547 GTNAGATCTGCCAGCTCCCGCTCGGCGGACGGCATCACCTCTGCGCGCGCGCCA 601

Search completed: March 13, 2005, 00:13:44  
Job time : 5214.18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 12, 2005, 16:41:47 ; Search time 261.349 Seconds  
(without alignments)  
9259.839 Million cell updates/sec

Title: US-09-475-704A-3  
Perfect score: 1479  
Sequence: 1 atggcgccgcgcgcacgacat.....acgacccctgagccagtaa 1479

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 segs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1197.8	81.0	1515	US-09-475-515-4	Sequence 4, Appli
2	1197.8	81.0	4472	US-09-475-515-75	Sequence 75, Appl
3	1197.8	81.0	4608	US-09-475-515-76	Sequence 76, Appl
4	1197.8	81.0	4689	US-09-475-515-74	Sequence 74, Appl
5	1197.8	81.0	4766	US-09-475-515-73	Sequence 73, Appl
6	1195.8	80.9	2031	US-09-475-515-7	Sequence 7, Appli
7	1114.6	75.4	1853	US-09-475-515-5	Sequence 5, Appli
8	1114.6	75.4	1865	US-09-475-515-78	Sequence 78, Appl
9	1114.6	75.4	1865	US-09-475-515-79	Sequence 79, Appl
10	1114.6	75.4	4319	US-09-475-515-6	Sequence 6, Appli
11	1025.2	69.3	1268	US-09-475-515-9	Sequence 9, Appli
12	976.6	66.0	1521	US-09-952-060-27	Sequence 27, Appl
13	976.6	66.0	1532	US-09-974-702-1	Sequence 1, Appli
14	976.6	66.0	1532	US-09-818-443-1	Sequence 11, Appli
15	976.6	66.0	37474	US-09-952-060-25	Sequence 25, Appl
16	973.6	65.8	4053	US-09-952-060-34	Sequence 34, Appl
17	949.4	64.2	4307	US-09-936-572-2	Sequence 2, Appli
18	947.6	64.1	4307	US-09-552-950-2	Sequence 2, Appli
19	939.8	63.5	4327	US-09-936-572-14	Sequence 14, Appl
20	939.8	63.5	4353	US-09-936-572-13	Sequence 13, Appl
21	939.8	63.5	4642	US-09-936-572-12	Sequence 12, Appl
22	939.8	63.5	9772	US-09-552-950-5	Sequence 5, Appli
23	933	63.1	1482	US-09-818-443-4	Sequence 4, Appli
24	930	62.9	1479	US-09-952-060-32	Sequence 32, Appl
25	744.6	50.3	8366	US-09-872-733A-6	Sequence 6, Appli
26	738.2	49.9	1476	US-09-184-418C-100	Sequence 100, App
27	738.2	49.9	8959	US-09-184-418C-11	Sequence 11, Appl

28	733.4	49.6	4338	4	US-09-872-733A-1	Sequence 1, Appli
29	731.4	49.5	9913	4	US-09-827-688-11	Sequence 11, Appl
30	729.8	49.3	1485	4	US-09-184-418C-73	Sequence 73, Appl
31	729.8	49.3	9010	4	US-09-184-418C-8	Sequence 8, Appli
32	728.4	49.2	12523	4	US-09-991-258-1	Sequence 1, Appli
33	726	49.1	1476	4	US-09-918-258-4	Sequence 38, Appl
34	710.8	48.1	1486	4	US-09-184-418C-38	Sequence 4, Appli
35	710.8	48.1	8992	4	US-09-184-418C-4	Sequence 9, Appli
36	701.2	47.4	1496	4	US-09-184-418C-82	Sequence 82, Appl
37	701.2	47.4	8972	4	US-09-184-418C-9	Sequence 29, Appl
38	659.8	44.6	1503	4	US-09-184-418C-29	Sequence 3, Appli
39	659.8	44.6	8953	4	US-09-184-418C-3	Sequence 12, Appl
40	658.8	44.5	1479	4	US-09-184-418C-12	Sequence 48, Appl
41	658.8	44.5	8968	4	US-09-184-418C-1	Sequence 1, Appli
42	656.6	44.4	4307	4	US-09-184-418C-48	Sequence 5, Appli
43	656.6	44.4	8966	4	US-09-184-418C-5	Sequence 47, Appl
44	652.6	44.1	1494	4	US-09-184-418C-47	Sequence 1, Appli
45	650.2	44.0	4307	4	US-09-552-950-1	

ALIGNMENTS

RESULT 1  
US-09-475-515-4  
; Sequence 4, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1515  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-Gag  
US-09-475-515-4

Query Match 81.0%; Score 1197.8; DB 4; Length 1515;  
Best Local Similarity 89.3%; Pred. No. 9.3e-177;  
Matches 1347; Conservative 0; Mismatches 132; Indels 30; Gaps 4;

QY	1	ATGGGCGCGCGCGCGCGCATCTGCGCGCGCGCGCAAGCTGAGCGGCTGGGAGCGCATCCG	60
DB	7	ATGGGCGCGCGCGCGCGCATCTGAGCGCGCGCGCAAGCTGAGCGGAGAGATCCG	66
QY	61	CTGGCGCGCGCGCGCGCAAGAGTGTACATGATGAACCACTGTGTGGGCGAGCGCGGAG	120
DB	67	CTGGCGCGCGCGCGCGCAAGAGTGTACATGATGAACCACTGTGTGGGCGAGCGCGGAG	126
QY	121	CTGGAGAGGTTGCGCTGAAACCGCGCGCTGCTGGAGACCAAGCGGCTGCAAGCATC	180
DB	127	CTGGAGCGCTTGGCTGTGAACCGCGCGCTGCTGGAGACCAAGCGGCTGCGCGCATC	186
QY	181	ATCCGCCAGCTGTGACCGCGCGCTGTGAGACCGGAGGAGGCTGAGAGCCTGTTCAC	240
DB	187	CTGGGCCAGCTGTGAGCGCGCGCTGTGAGACCGGAGGAGGAGGCTGTGCGCATCAAC	246

```
QY 241 ACCGTGGCCACCTGTACTGCGTGACGAGAAGATCGAGGTCCCGCACCAAGAGAGGCC 300
DB 247 ACCGTGGCCACCTGTACTGCGTGACGAGCGCATCGATCGTCAAGAGACCAAGAGAGGCC 306
QY 301 CTGGAACAAGATCGAGAGAGAGACAGAACAAGTGCAGACAGAAATCGACAGGCCGAGGCC 360
DB 307 CTGGAAGAAGATCGAGAGAGAGACAGAACAAGTCCAAAGAAAGGCCAGACAGGCCGCGGCC 366
QY 361 GCCG-----ACAAAGGCAAGGTGAGCCAGAACTACCCCATCTGTGACAGAACTTG 408
DB 367 GCCGCGGGCACCAGCAACAGACGCGGTGAGCCAGAACTACCCCATCTGTGACAGAACTTG 426
QY 409 CAGGCGCAGATGTGTGACACGAGCCATCAGCCCCCGCACCTGAACGCTGGGTGAAGGTG 468
DB 427 CAGGCGCAGATGTGTGACACGAGCCATCAGCCCCCGCACCTGAACGCTGGGTGAAGGTG 486
QY 469 ATGAGAGAGAGAGGCTTACGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGC 528
DB 487 GTGAGAGAGAGAGGCTTACGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGC 546
QY 529 GCCAACCCTCCAGGACTGAAACAGATGTGAACACCGTGGGCGGCCACAGGCCGCGCATG 588
DB 547 GCCAACCCTCCAGGACTGAAACAGATGTGAACACCGTGGGCGGCCACAGGCCGCGCATG 606
QY 589 CAGATGCTGAAGAGACCATCAACGAGAGAGGCCCGCAGTGGGACCGCGTGACCCCGTG 648
DB 607 CAGATGCTGAAGAGACCATCAACGAGAGAGGCCCGCAGTGGGACCGCGTGACCCCGTG 666
QY 649 CAGCGCGGCGCCCATCGCCCCCGGCGAGTGCAGAGCCCCGCGGACGACATCGCGGGC 708
DB 667 CAGCGCGGCGCCCATCGCCCCCGGCGAGTGCAGAGCCCCGCGGACGACATCGCGGGC 726
QY 709 ACCACCAAGACCTCGAGAGAGAGATGCGCTGATGACCAAGAAACCCCTCCATCCCGTG 768
DB 727 ACCACCAAGACCTCGAGAGAGAGATGCGCTGATGACCAAGAAACCCCTCCATCCCGTG 786
QY 769 GCGCATCTTACAAGCGGTGATCATCTTGGGCTGAAACAAGATCGTGGATGTACAGC 828
DB 787 GCGCATCTTACAAGCGGTGATCATCTTGGGCTGAAACAAGATCGTGGATGTACAGC 846
QY 829 CCGGTGAGCATCTGGAATCAACGAGAGAGGCCCAAGAGACCTTCCGCGACTACGTGAC 888
DB 847 CCGGTGAGCATCTGGAATCAACGAGAGAGGCCCAAGAGACCTTCCGCGACTACGTGAC 906
QY 889 CGCTTCTTCAAGACCTTGCAGCGCGCGAGAGACCAAGAGAGGTGAAGAACTGTGATGAC 948
DB 907 CGCTTCTTCAAGACCTTGCAGCGCGCGAGAGACCAAGAGAGGTGAAGAACTGTGATGAC 966
QY 949 GACACCTGCTGTGTGAGAAACGCAACCCGACTGCAAGACCATCTGCGCGCTCTGCGC 1008
DB 967 GACACCTGCTGTGTGAGAAACGCAACCCGACTGCAAGACCATCTGAGAGCTCTGCGC 1026
QY 1009 CCGGCGCGCAGCTGTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGGCCCGACAC 1068
DB 1027 CCGGCGCGCAGCTGTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGGCCCGACAC 1086
QY 1069 AAGGCGCGCGTGTGCGCGAGCGAGTGAAGCCAGGCCAAC-----CAGCGTGTGATG 1122
DB 1087 AAGGCGCGCGTGTGCGCGAGCGAGTGAAGCCAGGTGAAGCCAGGCCAACCATCATGATG 1146
QY 1123 CAGAAGAGCAACTTCAAGGGCCCGCGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAG 1182
DB 1147 CAGGCGCGCAACTTCCGCAACCAAGGGAAGACCGTCAAGTGTCTCAACTGCGGCAAGAG 1206
QY 1183 GCGCATCTGCGCGCAACTGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGCAAG 1242
DB 1207 GCGCATCTGCGCGCAAGAGTGTGCGCGCGCGCGCGCAAGAGGGCTGTGCGCGCGC 1266
QY 1243 GAGGCGCACCAAGATGAAGAGTGAACGAGCGCGCAAGGCCAATTCTGTGGCAAGATCTG 1302
DB 1267 GAGGCGCACCAAGATGAAGAGTGAACGAGCGCGCGCAAGGCCAATTCTGTGGCAAGATCTG 1326
QY 1303 CCGAGCCACAAGGCGCGCGCGCGCAACTTCTGCAAGAGCGCGCGCGAGCCACCGCGCC 1362
```

```
DB 1327 CCGAGTACAAGGGCGCGCGCGCAACTTCTGTGACAGAGCGCGCGCGAGCCCGCCCCC 1386
QY 1363 CCGCGCAGAGCTTCCGCTTC-----GAGGAGACCAACCCCGCGCAGAAAGAGAG--- 1413
DB 1387 CCGGAGAGAGCTTCCGCTTCCGCGCAGAGAGAGACCAACCCCGCAGAAAGAGAGGCC 1446
QY 1414 ---AGCAAGACCGCGAGACCTGACCAAGCTGAAGAGAGCTGTTCGCAACGACCCCTG 1470
DB 1447 ATGACAAGAGAGTGTATCCCGCTGACCAAGCTGCGCAGAGCTGTTCGCAACGACCCAGC 1506
QY 1471 AGCCAGTAA 1479
DB 1507 AGCCAGTAA 1515

RESULT 2
US-09-475-515-75
; Sequence 75, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 4472
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; US-09-475-515-75

Query Match      81.0%; Score 1197.8; DB 4; Length 4472;
Best Local Similarity 89.3%; Pred. No. 9.3e-177;
Matches 1347; Conservative 0; Mismatches 132; Indels 30; Gaps 4;

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGCGCAAGCTTGACGCTGGAGCGCATCCGC 60
DB 2939 ATGGGCGCGCGCGCAGCGTGTGAGCGCGCGAGCTGCAAGTGGAGAGATCCGC 2998
QY 61 CTGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGCACTGTGTGGCCAGCGCGAG 120
DB 2999 CTGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGCACTGTGTGGCCAGCGCGAG 3058
QY 121 CTGGAAGATTGCGCCCTGAACCCCGCGCTGTGAGAGCAAGCGAGGGCTGCAAGCATC 180
DB 3059 CTGGAAGCGTTGCGCCGTGAACCCCGCGCTGTGAGAGCAAGCGAGGGCTGCGCGCATC 3118
QY 181 ATCCGCAAGTGCACCCCGCGCTGACAGACCGCGCAGAGAGCTGAAGAGCTGTTCAC 240
DB 3119 CTGCGCGCAGCTGACCCAGCGCTGACAGACCGCGCAGAGAGCTGCGAGCTGTACAAC 3178
QY 241 ACCGTGGCCACCTGTACTGCGTGACGAGAAGATCGAGGTCCCGCACCAAGAGAGGCC 300
DB 3179 ACCGTGGCCACCTGTACTGCGTGACGAGCGCATCGACGTCAAGAGACCAAGAGAGGCC 3238
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTGCAGACAGAGATCTGAGAGCGCGAGGCC 360
DB 3239 CTGGAAGAAGATCGAGAGAGAGCAACAAGTCCAAAGAAAGGCCAGAGCGCGCGCC 3298
```



3495 CAGGGCCAGATGCTGACCAAGGCCATCAGCCCCCGCACCTTGAACGCTGGGTGAAGGTG 3554  
QY 469 ATGAGGAGAAAGGCTTCAAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGC 528  
DB 3555 GTGAGGAGAAAGGCTTCAAGCCCCGAGGTGATCCCATGTTCAAGCGCCCTGAGCGAGGGC 3614  
QY 529 GCCACCCCCAGAGACTGAACACAGATGTTGAACACCGTGGGGCGGCCACAGGGCCGCATG 588  
DB 3615 GCCACCCCCAGAGACTGAACACAGATGTTGAACACCGTGGGGCGGCCACAGGGCCGCATG 3674  
QY 589 CAGATGCTGAAGAGACCATCAACGAGAGAGGGCCCGGAGTGGAGCCGCTGCAACCCCGTG 648  
DB 3675 CAGATGCTGAAGAGACCATCAACGAGAGAGGGCCCGGAGTGGAGCCGCTGCAACCCCGTG 3734  
QY 649 CAGCGCGGGCCCATCGCCCCCGGCGAGATGCGCGAGCCCCCGGCGAGACATCGCCGGC 708  
DB 3735 CAGCGCGGGCCCATCGCCCCCGGCGAGATGCGCGAGCCCCCGGCGAGACATCGCCGGC 3794  
QY 709 ACCACCAGCACCCTGCAAGAGAGAGATCGCTGTGATGACCAAGCAACCCCCCATCCCGTG 768  
DB 3795 ACCACCAGCACCCTGCAAGAGAGAGATCGCTGTGATGACCAAGCAACCCCCCATCCCGTG 3854  
QY 769 GCGCATCTACAAAGCGGTGATCATCTGGGGCTGAAACAAGATCGTGGGATGTAAGC 828  
DB 3855 GCGCATCTACAAAGCGGTGATCATCTGGGGCTGAAACAAGATCGTGGGATGTAAGC 3914  
QY 829 CCCGTGAGCATCTGGAACATCAAGCAGGGCCCCCAAGAGCCCTTCGCGACTACGTGAGC 888  
DB 3915 CCCACCAAGCATCTGGAACATCGCGCAGAGGGCCCCCAAGAGCCCTTCGCGACTACGTGAGC 3974  
QY 889 CGCTTCTTCAAGACCTCTGCGCGCCGAGCAGAGACCCAGAGGTGAAGAACTGTGATGACC 948  
DB 3975 CGCTTCTTCAAGACCTCTGCGCGCCGAGCAGAGCCAGCCAGACGTGAAGAACTGTGATGACC 4034  
QY 949 GACACCTGTGTGTCAGAAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGC 1008  
DB 4035 GACACCTGTGTGTCAGAAACGCCAACCCCGACTGCAAGACCATCTGAGGCTCTCGGC 4094  
QY 1009 CCCGCGCCAGCCTGGAAGAGATGATGACCGCTGCGCAAGGCGGTGGCGGCCACAGCCAC 1068  
DB 4095 CCCGCGCCAGCCTGGAAGAGATGATGACCGCTGCGCAAGGCGGTGGCGGCCACAGCCAC 4154  
QY 1069 AAGGCCCGCGTGTGTCGAGGCGGATGAGCCAGGCCAACCA-----CAGCGTGTGATG 1122  
DB 4155 AAGGCCCGCGTGTGTCGAGGCGGATGAGCCAGGTGACGAACCCGCGCAGCATCATGTATG 4214  
QY 1123 CAGAAGAGCAACTTCAAGGGCCCCCGCGCATCTGTCAAGTGTCTTCAACTGCGGCAAGAG 1182  
DB 4215 CAGCGCGGCAACTTCCGCAACCAAGCGGAAAGACCTGTCAAGTGTCTTCAACTGCGGCAAGAG 4274  
QY 1183 GGCCCATGCGCCCGCAATGCGCGCGCCCCCGCAAGAAAGGCTGTGGAAGTGCAGCAAG 1242  
DB 4275 GGCCCATGCGCCCGCAATGCGCGCGCCCCCGCAAGAAAGGCTGTGCGCTGCGCGCGC 4334  
QY 1243 GAGGGCCACCAAGATGAAGAGACTGCAACGAGCGCGCAAGCCAACTTCTGGGCAAGATCTGG 1302  
DB 4335 GAGGGCCACCAAGATGAAGAGACTGCAACGAGCGCGCAAGCCAACTTCTGGGCAAGATCTGG 4394  
QY 1303 CCCAGCCACAAGGCGCGCCCCCGGCAACTTCTGTGAGAGCGCGCCCGAGCCACCGCCCC 1362  
DB 4395 CCCAGCTACAAGGCGCGCCCCCGGCAACTTCTGTGAGAGCGCGCCCGAGCCACCGCCCC 4454  
QY 1363 CCGCGCGAGAGCTTCGCTTC-----GAGGAGACCAACCCCGCGCAAGAGCAGAG--- 1413  
DB 4455 CCGGAGGAGAGCTTCGCTTCGCGGAGGAGAAAGACACCCCGAGCAAGAGCAGAGGCC 4514  
QY 1414 ---AGCAAGAGACCGGAGACCTGTGACCAAGCTGAAGAGCTGTTCGCAAGAGACCCCTG 1470  
DB 4515 ATGACCAAGAGAGCTGTACCCCTGTACCAAGCTGTGCGAGCTGTTCGCAAGAGACCCCAAGC 4574  
QY 1471 AGCCAGTAA 1479  
DB 4575 AGCCAGTAA 4583

RESULT 4  
US-09-475-515-74  
; Sequence 74, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475, 515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 74  
; LENGTH: 4689  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: sp160.modsFl62.gag.modsSF2  
US-09-475-515-74  
  
Query Match 81.0%; Score 1197.8; DB 4; Length 4689;  
Best Local Similarity 89.3%; Pred. No. 9.3e-177;  
Matches 1347; Conservative 0; Mismatches 132; Indels 30; Gaps 4;  
  
QY 1 ATGGGCGCCCGCGCCAGCATCTGCGCGCGCGCAAGCTGACGCTGGAGCGCATCCGC 60  
DB 3156 ATGGGCGCCCGCGCCAGCGTGTGAGCGCGCGCGAGCTGCAAGTGGAGAGATCCGC 3215  
QY 61 CTGCGCCCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGCGCCAGCGCGAG 120  
DB 3216 CTGCGCCCGCGCGCAAGAGAGTACAACTGAAAGCACATCTGTGGGCGAGCGCGAG 3275  
QY 121 CTGAGAAATTCGCTTGAACCCCGCGCTGTGAGAGCAGCGAGGGCTGCAAGCATC 180  
DB 3276 CTGAGAGCTTCGCGGTGAACCCCGCGCTGTGAGAGCAGCGAGGGCTGCGCCAGATC 3335  
QY 181 ATCCGCAAGTGCACCCCGCGCTGCAGACCGCGCAGCGAGAGCTGAAGAGCTGTTCAC 240  
DB 3336 CTGCGCCAGCTGCAGCCAGCTGCAGACCGCGCAGCGAGAGCTGCGAGCTGTTCAC 3395  
QY 241 ACCGTGCGCACTGTACTGTGTCAGCAGAGATCGAGTCCGCGACCAAGAGGCC 300  
DB 3396 ACCGTGCGCACTGTACTGTGTCAGCAGCGCATGCATGCAAGAGACCAAGAGGCC 3455  
QY 301 CTGACAAGATGAGAGAGAGAGCAACAAGTGCAGAGAGATCCAGCAGCGCGAGGCC 360  
DB 3456 CTGAGAAAGATGAGAGAGAGAGCAACAAGTGCAGAGAGGCCAGCAGCGCGCGC 3515  
QY 361 GCCG-----ACAAAGGCAAGGTGAGCCAGAACTAACCCCATCTGTGAGAACTGTG 408  
DB 3516 GCCGCGCGCACCGGCAACAGCAGCAGGTGAGCCAGAACTAACCCCATCTGTGAGAACTGTG 3575  
QY 409 CAGGGCCAGATGTGTCAACGAGCCATCAGCCCGCGCAACCTGAAGCGCTGGGTGAAGGTG 468  
DB 3576 CAGGGCCAGATGTGTCAACGAGCCATCAGCCCGCGCAACCTGAAGCGCTGGGTGAAGGTG 3635  
QY 469 ATGAGAGAGAGGCTTCAAGCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGC 528  
DB 3636 GTGAGAGAGAGGCTTCAAGCCCGAGGTGATCCCATGTTCAAGCGCCCTGAGCGAGGGC 3695  
QY 529 GCCAACCCTCAGAGACTGAACAGATGTTGAACACCGTGGCGGCCACCAAGCGCCCATG 588

```
DB 3696 GCCACCCCCCAGGAGCTGTAACAGATGTTGAACACCTGTGGGGCCCAAGGCCCGCATG 3755
QY 589 CAGATGCTGAAGGACCATCAACGAGGAGGGCCCGGAGTGGGACCGCGTGCAACCCCGTG 648
DB 3756 CAGATGCTGAAGGAGACCATCAACGAGGAGGGCCCGGAGTGGGACCGCGTGCAACCCCGTG 3815
QY 649 CAGCGCGGGCCCATCGCCCCCGGCGAGATGCGCGAGCGCCCGGCGAGCGACATCGCGGGC 708
DB 3816 CAGCGCGGGCCCATCGCCCCCGGCGAGATGCGCGAGCGCCCGGCGAGCGACATCGCGGGC 3875
QY 709 ACCACGACGACCCCTGCAAGAGAGAGATGCGCTGTGATGACGAAACCCCGCATCCCGTG 768
DB 3876 ACCACGACGACCCCTGCAAGAGAGAGATGCGCTGTGATGACGAAACCCCGCATCCCGTG 3935
QY 769 GCGCATCTCAAGCGGTGATCATCTGGGCTGTAAACAAGATCGTGGGATGTACAGC 828
DB 3936 GCGCATCTCAAGCGGTGATCATCTGGGCTGTAAACAAGATCGTGGGATGTACAGC 3995
QY 829 CCGGTGACATCTGTGACATCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 888
DB 3996 CCGGTGACATCTGTGACATCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4055
QY 889 CGCTTCTTCAAGACCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 948
DB 4056 CGCTTCTTCAAGACCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4115
QY 949 GACACCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1008
DB 4116 GACACCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4175
QY 1009 CCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1068
DB 4176 CCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4235
QY 1069 AAGGCGCGGCTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1122
DB 4236 AAGGCGCGGCTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4295
QY 1123 CAGAAAGAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1182
DB 4296 CAGCGCGGCAACTTCCGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4355
QY 1183 GGGCAATCGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242
DB 4356 GGGCAATCGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4415
QY 1243 GAGGGCCACAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302
DB 4416 GAGGGCCACAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4475
QY 1303 CCGAGCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1362
DB 4476 CCGAGCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4535
QY 1363 CCGCGCGAGAGCTTCCGCTT-----GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG-- 1413
DB 4536 CCGAGGAGAGAGCTTCCGCTTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4595
QY 1414 ---AGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG-- 1470
DB 4596 ATCGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4655
QY 1471 AGCCAGTAA 1479
DB 4656 AGCCAGTAA 4664
```

## RESULT 5

US-09-475-515-73  
; Sequence 73, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:

```
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
; LENGTH: 4766
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp160.modUS4.gag.modSF2
US-09-475-515-73
```

```
Query Match 81.0%; Score 1197.8; DB 4; Length 4766;
Best Local Similarity 89.3%; Pred. No. 9.3e-177;
Matches 1347; Conservative 0; Mismatches 132; Indels 30; Gaps 4;

QY 1 ATGGGGCGCGCGCGCAGCATCTGCGCGCGGCAAGCTGACGCGCTGGAGCGCATCCG 60
DB 3233 ATGGGGCGCGCGCGCAGCATCTGCGCGCGGCAAGCTGACGCGCGCTGGAGCAAGATCCG 3292
QY 61 CTGGCGCGCGCGCGCAAGAGTGTACATGATGAAGACCTGGTGGGCCAGCGCGAG 120
DB 3293 CTGGCGCGCGCGCGCAAGAGTGTACATGATGAAGACCTGGTGGGCCAGCGCGAG 3352
QY 121 CTGGAAGAGTGGCCCTGGAACCCCGCGCTGTGGAAGACCGAGGGGCTGCAAGAGATC 180
DB 3353 CTGGAAGAGTGGCCCTGGAACCCCGCGCTGTGGAAGACCGAGGGGCTGCGCGAGATC 3412
QY 181 ATCCGCCAGCTGACACCCCGCGCTGCAAGACCGGAGCGAGAGCTGAAGAGCTTTCAAC 240
DB 3413 CTGGGCGCAGCTGACACCCCGCGCTGCAAGACCGGAGCGAGAGCTGCGCGAGCTGTACAAC 3472
QY 241 ACCGTGGCCACCTGTACTGCTGTCAGAGAGATCGAGGTCCCGGACCAAGAGAGGCC 300
DB 3473 ACCGTGGCCACCTGTACTGCTGTCAGAGAGATCGAGGTCCCGGACCAAGAGAGGCC 3532
QY 301 CTGGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 3533 CTGGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3592
QY 361 GCCG-----ACAAGGCAAGGTGAGCGCAAGACTTACCCCATCTGTGAGAACCTG 408
DB 3593 GCCGCGCGCAGCGGCAACAGCAGCGAGGTGAGCGCAAGACTTACCCCATCTGTGAGAACCTG 3652
QY 409 CAGGGCCAGATGTGTACACAGGCGCATCAGCCCGGCAACCTGAACGCTGGGTGAAGGTG 468
DB 3653 CAGGGCCAGATGTGTACACAGGCGCATCAGCCCGGCAACCTGAACGCTGGGTGAAGGTG 3712
QY 469 ATCGAGGAGAGAGCTTCAAGCGCGAGGTATCCCATGTTCACCGCCCTGAGCGAGGGC 528
DB 3713 GTGAGGAGAGAGCTTCAAGCGCGAGGTATCCCATGTTCACCGCCCTGAGCGAGGGC 3772
QY 529 GCCACCCCGCAGAGACTGAACAGATGTTGAACACCGTGGGCGGCCACAGGCGGCATG 588
DB 3773 GCCACCCCGCAGAGACTGAACAGATGTTGAACACCGTGGGCGGCCACAGGCGGCATG 3832
QY 589 CAGATGCTGAAGAGACCATCAACGAGAGGCGCGGAGTGGAGCCGCTGACACCCCGTG 648
DB 3833 CAGATGCTGAAGAGACCATCAACGAGAGGCGCGGAGTGGAGCCGCTGACACCCCGTG 3892
```

```
QY 649 CAGCGCGCGCCATCGCCCCCGGCGAGATGCGCGAGCCCCCGGCGAGCGACATCGCCGGC 708
    |||
DB 3893 CAGCGCGCGCCATCGCCCCCGGCGAGATGCGCGAGCCCCCGGCGAGCGACATCGCCGGC 3952
QY 709 ACCACCGAGACCTTGAGAGAGAGATGCGCTGTGATGACCAAGCAACCCCCCATCCCGGTG 768
    |||
DB 3953 ACCACCGAGACCTTGAGAGAGAGATGCGCTGTGATGACCAAGCAACCCCCCATCCCGGTG 4012
QY 769 GCGGACATCTACAAGCGGTGATCATCTGGGCGCTGAACAAGATGCTGCGGATGTACAGC 828
    |||
DB 4013 GCGGAGATCTACAAGCGGTGATCATCTGGGCGCTGAACAAGATGCTGCGGATGTACAGC 4072
QY 829 CCGGTGAGCATCTTGAACATCAAGAGGGCCCCCAAGAGACCTTCCGCGACTACGTGAGC 888
    |||
DB 4073 CCGACCGACATCTTGAACATCGCGCGAGGGCCCCCAAGAGACCTTCCGCGACTACGTGAGC 4132
QY 889 CGCTTCTTCAAGACCTTGGCGCGCGAGAGAGACCCAGAGAGTGAAGAACTGATGACC 948
    |||
DB 4133 CGCTTCTTCAAGACCTTGGCGCGCGAGAGAGACCCAGAGAGTGAAGAACTGATGACC 4192
QY 949 GACACCTTGTGTGAGAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGC 1008
    |||
DB 4193 GAGACCTTGTGTGAGAGAACGCCAACCCCGACTGCAAGACCATCTGAGGCTCTCGGC 4252
QY 1009 CCGGCGCGAGCCTTGAGAGAGATGATGACCGCTGCGAGGGCGTGGGCGGCCAGCCAC 1068
    |||
DB 4253 CCGGCGCGAGCCTTGAGAGAGATGATGACCGCTGCGAGGGCGTGGGCGGCCAGCCAC 4312
QY 1069 AAGGCGCGGTGTGCGCGAGGCGATGAGCCAGGCCAACCA-----CGACGTGATGATG 1122
    |||
DB 4313 AAGGCGCGGTGTGCGCGAGGCGATGAGCCAGGTGAGCAACCGGCGAGCATCATGATG 4372
QY 1123 CAGAAAGCAACTTCAAGGGCGCGCGCGCATCTCAAGTGTCTTCAACTGCGGCAAGAG 1182
    |||
DB 4373 CAGCGCGGCAACTTCCGCAACCAAGCGGAGAGACCTTCAAGTGTCTTCAACTGCGGCAAGAG 4432
QY 1183 GCGCACATCGCCCGCAACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGGCAAG 1242
    |||
DB 4433 GCGCACATCGCCCGCAACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGGCGCGC 4492
QY 1243 GAGGGCCACCAAGATGAAGGACTGACCGAGCGCGCAAGCTTCTGGGCAAGATCTGG 1302
    |||
DB 4493 GAGGGCCACCAAGATGAAGGACTGACCGAGCGCGCAAGCTTCTGGGCAAGATCTGG 4552
QY 1303 CCGAGCCACAAGGGCGCGCGCGCAACTTCTGTGAGAGCGCGCGCGAGCCCGCGCGCC 1362
    |||
DB 4553 CCGAGCTACAAGGGCGCGCGCGCGCAACTTCTGTGAGAGCGCGCGCGAGCCCGCGCC 4612
QY 1363 CCGCGCGAGAGCTTCCGCTTC-----GAGAGACCACCCCCGCGCAAGAGCAGAG--- 1413
    |||
DB 4613 CCGGAGAGAGCTTCCGCTTGGCGAGAGAGACCACCCCAGCGCAAGAGCAGAGCCCC 4672
QY 1414 ---AGCAAGAGACCGGAGACCTGACCGCTGAGAGCGCTGTTCGCAACGAGCCCGCTG 1470
    |||
DB 4673 ATGCAACAAGAGAGCTGTACCCCTGTGACCAAGCTGTGCGAGCGCTGTTCGCAACGAGCCCGAGC 4732
QY 1471 AGCCAGTAA 1479
    |||
DB 4733 AGCCAGTAA 4741
    |||
```

## RESULT 6

```
US-09-475-515-7
; Sequence 7, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
```

```
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: HIV-Gag/HCV-core fusion polypeptide
US-09-475-515-7
```

```
Query Match 80.9%; Score 1195.8; DB 4; Length 2031;
Best Local Similarity 89.3%; Pred. No. 1.9e-176;
Matches 1345; Conservative 0; Mismatches 132; Indels 30; Gaps 4;
```

```
QY 1 ATGGGCGCGCGCGCGAGCATCTGCGCGGCGCAAGCTGAGCGCTGGAGCGCATCCG 60
    |||
DB 7 ATGGGCGCGCGCGCGAGCATCTGCGCGGCGCAAGCTGAGCGCGGCGCAAGCTGAGCGCATCCG 66
QY 61 CTGCGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGCACCCTGTGTGGGCCAGCGCGAG 120
    |||
DB 67 CTGCGCGCGCGCGCGCAAGAGAGTGAAGCTGAAGCACAATGTGTGGGCCAGCGCGAG 126
QY 121 CTGAGAAAGTTTCCCTTGAACCCCGCTGCTGAGAGACCAAGAGGCTGCAACAGATC 180
    |||
DB 127 CTGAGCGCTTTCGCGGTGAACCCCGCTGCTGAGAGACCAAGAGGCTGCGCGAGATC 186
QY 181 ATCCGCGAGCTGCAACCCCGCTGCAAGACCGCGAGAGAGAGCTGAAGAGCTGTCAAC 240
    |||
DB 187 CTGGGCGAGCTGCAACCCCGCTGCAAGACCGCGAGAGAGAGCTGCGAGCTGTACAC 246
QY 241 ACCGTGGCAACCTGTACTGTGCTGCAAGAGATCGAGTTCGCGACACCAAGAGGCC 300
    |||
DB 247 ACCGTGGCAACCTGTACTGTGCTGCAAGAGATCGAGTTCGCGACACCAAGAGGCC 306
QY 301 CTGACAAAGATCGAGGAGAGAGCAAGTGCAGAGAAATCCAGAGAGCGCGAGGCC 360
    |||
DB 307 CTGAGAAAGATCGAGGAGAGAGCAAGTGCAGAGAAAGGCCAGAGAGCGCGGCC 366
QY 361 GCCG-----ACAAAGGCAAGGTGAGCCAGAACTAACCCCATCTGTGCAAACTTG 408
    |||
DB 367 GCCGCGCGCAACCGCAACAGAGCCAGGTGAGCCAGAACTAACCCCATCTGTGCAAACTTG 426
QY 409 CAGGCGCAAGTGTGCAACCAAGGCATCAAGCCCCCGCAACCTTGAACGCTGGTGAAGTG 468
    |||
DB 427 CAGGCGCAAGTGTGCAACCAAGGCATCAAGCCCCCGCAACCTTGAACGCTGGTGAAGTG 486
QY 469 ATGAGAGAGAGGCTTCAAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGC 528
    |||
DB 487 GTGAGAGAGAGGCTTCAAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGGC 546
QY 529 GCCACCCCCCAGAGACTTGAACAAGATGTTGAACAACGTTGGGCGGCCACAGGCGGCATG 588
    |||
DB 547 GCCACCCCCCAGAGACTTGAACAAGATGTTGAACAACGTTGGGCGGCCACAGGCGGCATG 606
QY 589 CAGATGCTGAAGAGACCATCAACGAGAGGCGCGGAGTGGGACCGCGTGCACCCCGTG 648
    |||
DB 607 CAGATGCTGAAGAGACCATCAACGAGAGGCGCGGAGTGGGACCGCGTGCACCCCGTG 666
QY 649 CAGCGCGCGCCATCGCCCCCGCGAGATGCGCGAGCCCGCGAGCGAGCATCGCGGC 708
    |||
DB 667 CAGCGCGCGCCATCGCCCCCGCGAGATGCGCGAGCCCGCGAGCGAGCATCGCGGC 726
QY 709 ACCACGAGACCTTGAGAGAGAGATCGCTGTGATGACCAAGCAACCCCCCATCCCGGTG 768
    |||
DB 727 ACCACGAGACCTTGAGAGAGAGATCGCTGTGATGACCAAGCAACCCCCCATCCCGGTG 786
```

QY	769	GGCGCATCTCAAGCGGTGGATCATCTTGGGCGCTGAACCAAGATCGTGGGATGTACAGC	828
Db	787	GGCGAGATCTTCAAGCGGTGGATCATCTTGGGCGCTGAACCAAGATCGTGGGATGTACAGC	846
QY	829	CCCGTGAGCATCTCTGACATCAAGCAGGGGCCCAAGAGACCCCTTCGCGACTACGTGAC	888
Db	847	CCCAACGACATCTCTGACATCCGCGCAGGGGCCCAAGAGACCCCTTCGCGACTACGTGAC	906
QY	889	CGCTTCTTCAAGACCCCTGGCGCGCCGACGACGCCACCAAGAGGTGAAGACTGGATGACC	948
Db	907	CGCTTCTTCAAGACCCCTGGCGCGCTGACGACGCCACCAAGAGGTGAAGACTGGATGACC	966
QY	949	GACACCCCTGCTGGTGCAGAACGCCCAACCCCGACTGCAAGACCACTCTGCGCGCTCTGGC	1008
Db	967	GACACCCCTGCTGGTGCAGAACGCCCAACCCCGACTGCAAGACCACTCTGGAAGGCTCTGGC	1026
QY	1009	CCCCGGCCGACGCTGGAGAGAGATGATACCGCTTCCAGGGCGTGGGGCGGCCCAAGCCAC	1068
Db	1027	CCCCGGCCGACCCCTGGAGAGAGATGATACCGCTTCCAGGGCGTGGGGCGGCCCAAGCCAC	1086
QY	1069	AAGGCCCCGCTGCTGGCGCCGAGGCGCATGAGCCAGCCACA-----CCAGCGTATGATG	1122
Db	1087	AAGGCCCCGCTGCTGGCGCCGAGGCGCATGAGCCAGGTGACGAACCGCGCATCATGATG	1146
QY	1123	CAGAAGAGCAACTTCAAGGGCCCCCGCGCATGCTCAAGTGTCTTCACTGCGGCAAGAG	1182
Db	1147	CAGCGCGGCAACTTCCGCAACCAAGCGGAAGACCGTCAAGTGTCTTCACTGCGGCAAGAG	1206
QY	1183	GGCCACATCGCCCGCAACTGCGCGCCCCCGGCAAGAGGCGTGTGGAAGTGGCGCAAG	1242
Db	1207	GGCCACACCGCCAGAAACTGCGCGCCCCCGGCAAGAGGCGTGTGGAAGTGGCGCGC	1266
QY	1243	GAGGGCCACCAAGATGAAGGACTGCACCGAGCGCCAGGCCCACTTCTTGGAAGATCTGG	1302
Db	1267	GAGGGCCACCAAGATGAAGGACTGCACCGAGCGCCAGGCCCACTTCTTGGAAGATCTGG	1326
QY	1303	CCCAAGCACAAGGGCGCGCCCGGCAACTTCTTGCAAGCGCGCCCGGAGCCACCGCCCCC	1362
Db	1327	CCCAAGCACAAGGGCGCGCCCGGCAACTTCTTGCAAGCGCGCCCGGAGCCACCGCCCCC	1386
QY	1363	CCCGCCGAGAGCTTCCGCTTC-----GAGGAGACCAACCCCGGCGAGAGCAGAG---	1413
Db	1387	CCCGAGGAGAGCTTCCGCTTCCGCGAGGAGAGACCAACCCCGGCGAGAGCAGAGGCC	1446
QY	1414	---AGCAAGGACCGCGAGAGACCTGACCAAGCCTGAAGAGCCTGTTCGGCAACGACCCCTG	1470
Db	1447	ATCGACAAAGAGCTGTACCCCTGACCAAGCCTGCGCAGCCTGTTCGGCAACGACCCCGAGC	1506
QY	1471	AGCCAGT 1477	
Db	1507	AGCCAGT 1513	

RESULT 7  
US-09-475-515-5  
; Sequence 5, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475, 515A  
; CURRENT FILING DATE: 1999-12-30

```

; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1853
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: HIV-Gag-protease
US-09-475-515-5

```

Query Match	75.4%;	Score 1114.6;	DB 4;	Length 1853;
Best Local Similarity	85.8%;	Pred. No. 6.6e-164;		
Matches 1295; Conservative	0;	Mismatches 184;	Indels 30;	Gaps 4;

QY	1	ATGGGCGCCCGGCCAGCATCTCTGCGCGGGCGGCAAGCTGGACGCTGTGGAGCGCATCCGC	60
Db	7	ATGGGCGCCCGGCCAGCGTGTGAGCGGGCGGAGCTGGACAAGTGTGGAGAAAGATCCGC	66
QY	61	CTGCGCCCCGGCGGCAAGAGTGTCTACATGATGAAGCACTGTGTGTGGGCCAGCCGCGAG	120
Db	67	CTGCGCCCCGGCGGCAAGAGAGTACAAAGCTGAAGCAATCTGTGTGGGCCAGCCGCGAG	126
QY	121	CTGGAAGAAGTTGCGCCCTGAAACCCCGGCGCTGCTGAGAACCAAGCGAGGCTGCAGCAGATC	180
Db	127	CTGGAAGCGCTTGCGCCGTGAACCCCGGCGCTGCTGAGAACCAAGCGAGGCTGCAGCAGATC	186
QY	181	ATCCGCCAGCTGCAACCCCGGCGCTGCAGACCCGCGAGGAGCTGAAGACCTGTTCAAC	240
Db	187	CTGGGCCAGCTGCAACCCCGGCGCTGCAGACCCGCGAGGAGCTGTGCGCAGCCTGTACAC	246
QY	241	ACCGTGGCCACCTGTACTGCGTGCAAGAAAGATCGAGGTCCGCGACACCAAGAGGCGC	300
Db	247	ACCGTGGCCACCTGTACTGCGTGCAAGAAAGATCGAGGTCCGCGACACCAAGAGGCGC	306
QY	301	CTGACAAGATCGAGGAGGAGCAAAAGTGCAGCAGAAAGATCCAGCAGGCGGAGCGC	360
Db	307	CTGAGAAGATCGAGGAGGAGCAAAAGTCCAAAGAAAGAGGCCAGCAGGCGGCGCGC	366
QY	361	GCCG-----ACAAAGGCAAGGTGAGGCCAGAACTACCCCATCTGTGCAAACTG	408
Db	367	GCCGCGGCAACGGCAACAGAGCCAGGTGAGCCAGAACTACCCCATCTGTGCAAACTG	426
QY	409	CAGGGCAGATGCTGCAACCAAGCCATCAGCCCCCGCACCTGAAAGCTGTGGTGAAGTG	468
Db	427	CAGGGCAGATGCTGCAACCAAGCCATCAGCCCCCGCACCTGAAAGCTGTGGTGAAGTG	486
QY	469	ATCGAGAGAAGGACCTTCAGCCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGC	528
Db	487	GTGAGAGAGAAGGACCTTCAGCCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGC	546
QY	529	GCCACCCCCCAGGACCTGAAACAGATGTTGAACACCGTGGCGGCAACAGGCGCGCATG	588
Db	547	GCCACCCCCCAGGACCTGAAACAGATGTTGAACACCGTGGCGGCAACAGGCGCGCATG	606
QY	589	CAGATGCTGAAGGACCACTCAACGAGAGGCGCGGAGTGGGACCGCGTGACCCCGTG	648
Db	607	CAGATGCTGAAGGACCACTCAACGAGAGGCGCGGAGTGGGACCGCGTGACCCCGTG	666
QY	649	CAGCGCGGCGCCATCGCCCCCGGCGAGATGCGCGAGCCCCGCGGAGGACATCGCCGGC	708
Db	667	CAGCGCGGCGCCATCGCCCCCGGCGAGATGCGCGAGCCCCGCGGAGGACATCGCCGGC	726
QY	709	ACCACCAAGACCTGCAAGAGCAGATCGCTGGATGACCAAGACCCCCCATCCCGTG	768
Db	727	ACCACCAAGACCTGCAAGAGCAGATCGCTGGATGACCAAGACCCCCCATCCCGTG	786
QY	769	GGCGACATCTACAAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGATGTACAGC	828
Db	787	GGCGACATCTACAAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGATGTACAGC	846
QY	829	CCCGTAGCATCTTGACATCAAGCAGGGCCCCAAGAGCCCTTCCGCGCTACGTGAC	888

```
DB 847 CCCACGACATCTGGACATCCGCCAGGGCCCCAAGAGAGCCCTTCGCGACTACGTGAC 906
QY 889 CGCTTCTTCAAGACCTGCGCGCCGAGCAGACCCAGAGGTGAAGAACTGGATGACC 948
DB 907 CGCTTCTACAGACCTGCGCGCTGAGCAGGCCAGCAGACCTGAAGAACTGGATGACC 966
QY 949 GACACCTGCTGTGAGAGCCCAACCCCGACTGCAAGACCACTCTGCGCGCTCTCGGC 1008
DB 967 GAGACCTGCTGTGAGAGCCCAACCCCGACTGCAAGACCACTCTGAGGCTCTCGGC 1026
QY 1009 CCGCGCGCAGCTTGAGAGAGATGATGACCGCTGCGAGGCGCGCGGCCAGCCAC 1068
DB 1027 CCGCGCGCAGCTTGAGAGAGATGATGACCGCTGCGAGGCGCGCGGCCAGCCAC 1086
QY 1069 AAGCGCGCGCTGCGCGAGCGAGTGAAGCCCAACA-----CCAGCGTGTATG 1122
DB 1087 AAGCGCGCGCTGCGCGAGCGAGTGAAGCCCAACA-----CCAGCGTGTATG 1146
QY 1123 CAGAAAGCAACTTCAAGGCGCGCGCGCATCGTCAAGTGTCTCAACTGCGCAAGAG 1182
DB 1147 CAGCGCGCAACTTCCGCAACCGAGCGAAGACCGTCAAGTGTCTCAACTGCGCAAGAG 1206
QY 1183 GCGCAACATGCGCGCAACTGCGCGCGCGCGCATCGTCAAGTGTCTGGAAGTGGCGCAAG 1242
DB 1207 GCGCAACATGCGCGCAAGACTGCGCGCGCGCGCGCAAGAGGCTGCGCGCGCGC 1266
QY 1243 GAGGGCCACCAAGATGAAGACTGCAAGCGCGCAAGCGCAACTCTCTGGCAAGATCTGG 1302
DB 1267 GAAGGACACCAATGAAGATGCACTGAGACAGGCTAATTTTAGGAAGATCTGG 1326
QY 1303 CCAGGCCCAAGGGCGCGCGCGCAACTCTCTGAGAGCGCGCGCGCGCGCGCGCGCG 1362
DB 1327 CCTTCTTACAAGGGAAGGCGCAGGGAATTTCTTCAAGACAGACCGCAAGCGCGCGCA 1386
QY 1363 CCGCGCGAGAGCTTCCGCTTGAAGAG-----ACCAACCGCGCGCGCAAGCAGAG--- 1413
DB 1387 CCAGAAGAGAGCTTCAAGTTGGGAGAGAGAAACAACCTCTCTCAAGAAGCAGAGCGG 1446
QY 1414 ---AGCAAGAGCCGCGAGACCTGACAGCTGAGAGAGCTGTCGCAACGACCCCTG 1470
DB 1447 ATAGACAAGAACTGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCG 1506
QY 1471 AGCCAGTAA 1479
DB 1507 TCACAGTAA 1515
```

```
RESULT 8
US-09-475-515-78
; Sequence 78, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 1865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

/ OTHER INFORMATION: Description of Artificial Sequence: GPI  
US-09-475-515-78

Query Match 75.4%; Score 114.6; DB 4; Length 1865;  
Best Local Similarity 85.8%; Pred. No. 6.6e-164;  
Matches 1295; Conservative 0; Mismatches 184; Indels 30; Gaps 4;

```
QY 1 ATGGGCGCGCGCGCAGCATCTCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60
DB 13 ATGGGCGCGCGCGCAGCGGCTGAGCGCGCGCGAGCTGAGCAAGTGGAGAAAGATCCGC 72
QY 61 CTGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGCACTGTGTGGCGCAGCGCGAG 120
DB 73 CTGCGCGCGCGCGCAAGAGATGATCAAGCTGAAGCACTGTGTGGCGCAGCGCGAG 132
QY 121 CTGGAAGATTCGCGCTGAACCCCGCGCTGAGAGCCAGCGAGGCTGCAAGCATC 180
DB 133 CTGGAAGCGCTTCGCGCTGAACCCCGCGCTGAGAGCCAGCGAGGCTGCGCGCATC 192
QY 181 ATCCGCGCAGTGCACCCCGCGCTGCAAGACCGCGCAGGAGCTGAAGAGCTTCAAC 240
DB 193 CTGGGCGCAGCTGCAGCGCGCGCTGCAAGACCGCGCAGGAGAGCTGCGCGCTGTACAAC 252
QY 241 ACCGTGGCCACCTGTACTGTGCTGCAAGAGATCGAGTCCGCGCACCAAGAGAGGCC 300
DB 253 ACCGTGGCCACCTGTACTGTGCTGCAAGACCGCGCATCGACTCAAGAGACCAAGAGAGGCC 312
QY 301 CTGGAACAAGTTCAGAGAGGAGCAGAACAGTCCAGCAGAGATCCAGCAGCGCGAGGCC 360
DB 313 CTGGAAGAAGTTCAGAGAGGAGCAGAACAGTCCAGAGAGAGGCCAGCAGCGCGCGCC 372
QY 361 GCCG-----ACAAGGGCAAGGTGAGCCAGAACTAACCCCATCTGTGAGAACCTG 408
DB 373 GCCGCGCGCGCAGCGCGCAACAGCAGCGCAGGTGAGCCAGAACTAACCCCATCTGTGAGAACCTG 432
QY 409 CAGGGCCAGATGTGCAACAGCGCCATCAAGCGCGCGCAGCCCTGAACGCGCTGGTGAAGTG 468
DB 433 CAGGGCCAGATGTGCAACAGCGCCATCAAGCGCGCGCAGCCCTGAACGCGCTGGTGAAGTG 492
QY 469 ATCGAGAGAGAGGCTTCAAGCGCGCGAGGTATCCCATGTTCAACGCGCTGAGCGAGGGC 528
DB 493 GTGAGAGAGAGAGGCTTCAAGCGCGCGAGGTATCCCATGTTCAAGCGCGCTGAGCGAGGGC 552
QY 529 GCCAAGCGCGCGCAGCCTGAACAAGATGTTGAACAACCGTGGCGCGCAGCAGCGCGCATG 588
DB 553 GCCAAGCGCGCGCAGCCTGAACAAGATGTTGAACAACCGTGGCGCGCAGCAGCGCGCATG 612
QY 589 CAGATGCTGAAGAGACACCATCAACAGAGAGCGCGCGAGTGGGAGACCGCGTGAACCCCGTG 648
DB 613 CAGATGCTGAAGAGAGACCATCAACAGAGAGCGCGCGAGTGGGAGACCGCGTGAACCCCGTG 672
QY 649 CAGCGCGCGCGCATCGCGCGCGCGCAGATGCGCGAGCGCGCGCGCGCGCAGCATCGCGCGC 708
DB 673 CAGCGCGCGCGCATCGCGCGCGCGCAGATGCGCGAGCGCGCGCGCGCGCAGCATCGCGCGC 732
QY 709 ACCACCAAGACCTGCAAGAGCAGATCGCGCTGGATGACCAAGCAACCCCGCATCCCGGTG 768
DB 733 ACCACCAAGACCTGCAAGAGCAGATCGCGCTGGATGACCAAGCAACCCCGCATCCCGGTG 792
QY 769 GCGGACATCTAACAAGCGGTGATCATCTGGGCTTGAACAAGATCGTGGGATGTACAGC 828
DB 793 GCGGACATCTAACAAGCGGTGATCATCTGGGCTTGAACAAGATCGTGGGATGTACAGC 852
QY 829 CCGGTGAGCATCTGACATCAAGCAGGCGCGCAAGAGCGCTTCCGCGACTACGTGAGC 888
DB 853 CCGGTGAGCATCTGACATCAAGCAGGCGCGCAAGAGCGCTTCCGCGACTACGTGAGC 912
QY 889 CGCTTCTTCAAGACCTGCGCGCGCGCAGAGCAGCAGCAGAGGTGAAGAACTGGATGACC 948
DB 913 CGCTTCTTCAAGACCTGCGCGCGCGCAGAGCAGCAGCAGAGGTGAAGAACTGGATGACC 972
QY 949 GACACCTGCTGTGAGAGCCCAACCCCGACTGCAAGACCACTCTGCGCGCTCTCGGC 1008
```

Db 973 GAGACCTGCTGTGTGAGAAAGCCAAACCCGACTGTGCAAGACCATCTGAAGGCTTCGGGC 1032  
QY 1009 CCCGGCGCCAGCCTGGAGAGATGATGACCGCCTGCCAGGGCGTGGCGGGCCAGCCAC 1068  
Db 1033 CCCGGCGCCAGCCTGGAGAGATGATGACCGCCTGCCAGGGCGTGGCGGGCCAGCCAC 1092  
QY 1069 AAGGCCCGCGTGTGGCCGAGGCGATGAGCCAGGCCAACCA-----CCAGCGTGATGATG 1122  
Db 1093 AAGGCCCGCGTGTGGCCGAGGCGATGAGCCAGGTGACGAACCCGGGACCATCATGATG 1152  
QY 1123 CAGAAGACCAACTTCAAGGGGCCCCCGCGCATCTGTCAAGTGTCTTCACTGCGGCAAGAG 1182  
Db 1153 CAGCGCGGCAACTTCCGCAACCAAGCGGAAGACCGTCAAGTGTCTTCACTGCGGCAAGAG 1212  
QY 1183 GGCCCATCGCCCGCACTGCGCGCGCGCGCGCGCGCAAGAGGGCTGTGGAAGTGGCGCAAG 1242  
Db 1213 GGCCCATCGCCCGCAAGTGTGCGCGCGCGCGCGCGCGCAAGAGGGCTGTGCGCGCGC 1272  
QY 1243 GAGGCGCACCATGAAAGGACTGCAACGAGCGCGCAAGCCAACTTCTGGGCAAGATCTGG 1302  
Db 1273 GAAGGACACCAATGAAGAATTGCACTGAGAGACAGGCTAATTTTAAAGGAAGATCTGG 1332  
QY 1303 CCCAGCCACAAGGCGCGCGCGCGCAACTTCTGCAAGAGCGCGCGCGCAAGCCAGCCCGCC 1362  
Db 1333 CTTCTCTACAAGGGAAGGCGAGGAATTTCTTCAAGACAGACCAAGCCAAAGCCCGCA 1392  
QY 1363 CCCGCGAGAGCTTCCGCTTCGAGAG-----ACCAACCCCGCGCGCAAGCAGAG--- 1413  
Db 1393 CCAGAAGAGAGCTTCAAGTTTGAGGAGAGAAACAACCTCCCTCTGAGAAAGCAGAGCCG 1452  
QY 1414 ---AGCAAGAGACCGGAGACCTGACCAAGCTGAAAGCCTGTTGGCAACGACCCCTG 1470  
Db 1453 ATAGACAAGAACTGTATCTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCG 1512  
QY 1471 AGCCAGTAA 1479  
Db 1513 TCACAGTAA 1521

RESULT 9  
US-09-475-515-79  
; Sequence 79, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475, 515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 79  
; LENGTH: 1865  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GP2  
US-09-475-515-79

Query Match 75.4%; Score 1114.6; DB 4; Length 1865;  
Best Local Similarity 85.8%; Pred. No. 6.6e-164;  
Matches 1295; Conservative 0; Mismatches 184; Indels 30; Gaps 4;

QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGCGCAAGCTGGAACGCTTGGGAGCGCATCCGC 60

Db 13 ATGGGCGCGCGCGCAGCGTGTGAGCGGCGGAGCTGACAAAGTGGAGAAAGATCCGC 72  
QY 61 CTGGCCCCCGGCGCAAGAAGTCTACATGATGAAGCACTGTGTGGCCAGCCCGAG 120  
Db 73 CTGGCCCCCGGCGCAAGAAGTACAAAGTGAAGCACATCGTGTGGCCAGCCCGAG 132  
QY 121 CTGGAAGATTGCGCTTGAACCCCGGCTGTGAGAGACCAAGAGGGCTGCAAGCAGATC 180  
Db 133 CTGAGCGCTTGCCTGTGAACCCCGGCTGTGAGAGACCAAGAGGGCTGCCAGATC 192  
QY 181 ATCCGCGAGCTGCAACCCCGCTGCAAGCCGCGAGGAGAGTGAAGACTGTCAAC 240  
Db 193 CTGGGCCAGCTGCAAGCCCGCTGCAAGCCGCGAGGAGAGCTGCCAGCTGTACAAC 252  
QY 241 ACCGTGGCCACCTGTACTGTGTCACGAGAAGATCGAGTCCGCGACCAAGAGGCC 300  
Db 253 ACCGTGGCCACCTGTACTGTGTCACGAGCGCATCGACGTCAAGACACCAAGAGGCC 312  
QY 301 CTGGAACAAGATCGAGAGAGAGCAACAAGTGCAGCAAGATCCAGAGGCCAGGCC 360  
Db 313 CTGGAAGAAGATCGAGAGAGAGCAACAAGTCCAAAGAAGAGGCCAGAGGCCCGCC 372  
QY 361 GCGG-----ACAAGGCAAGGTGAGCCAGACTACCCCATCTGTGCAAGACTG 408  
Db 373 GCGCGCGGCAACCGCAACAGCAGCCAGGTGAGCCAGACTACCCCATCTGTGCAAGACTG 432  
QY 409 CAGGCGCAGATGTGTGCAACAGGCCATCAGCCCGCGCACCTGAAACGCTGGGTGAAGTG 468  
Db 433 CAGGCGCAGATGTGTGCAACAGGCCATCAGCCCGCGCACCTGAAACGCTGGGTGAAGTG 492  
QY 469 ATGAGAGAGAGGCTTCAAGCCCGGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGC 528  
Db 493 GTGAGAGAGAGGCTTCAAGCCCGGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGC 552  
QY 529 GCCACCCCGCAGGACCTGAAACAGATGTGAACACCGTGGCGGCCACCAAGCGCCATG 588  
Db 553 GCCACCCCGCAGGACCTGAAACAGATGTGAACACCGTGGCGGCCACCAAGCGCCATG 612  
QY 589 CAGATGCTGAAGGACACCATCAACGAGAGGCGCGGAGTGGACCGCGTGCACCCCGTG 648  
Db 613 CAGATGCTGAAGGAGACCATCAACGAGAGGCGCGGAGTGGACCGCGTGCACCCCGTG 672  
QY 649 CAGCGCGCGCCCATGCGCCCCCGGCGAGATGCGCGAGCCCGCGGAGCGACATCGCGGC 708  
Db 673 CAGCGCGCGCCCATGCGCCCCCGGCGAGATGCGCGAGCCCGCGGAGCGACATCGCGGC 732  
QY 709 ACCACGAGCACTCTGCAAGAGAGAGATGCGCTGATGACCAAGCAACCCCGCATCCCGTG 768  
Db 733 ACCACGAGCACTCTGCAAGAGAGAGATGCGCTGATGACCAAGCAACCCCGCATCCCGTG 792  
QY 769 GCGCATCTTACAAGCGGTGATCATCTGGGCGCTGAAACAAGATCGTGGATGTACAGC 828  
Db 793 GCGAGATCTACAAGCGGTGATCATCTGGGCGCTGAAACAAGATCGTGGATGTACAGC 852  
QY 829 CCCGTGAGCATCTGGAACATCAAGAGGCGCCCAAGAGGCCCTTCCGCGACTACGTGAC 888  
Db 853 CCCACGAGCATCTGGAACATCGCGCAGGCGCCCAAGAGGCCCTTCCGCGACTACGTGAC 912  
QY 889 CGCTTCTTACAAGACCTTGCAGCGCGGAGAGAGCAACCGAGCTGCAAGACCATCTGCGCTCGGC 948  
Db 913 CGCTTCTTACAAGACCTTGCAGCGCGTGAAGAGCAAGCGAGCGTGAAGAACTGTGAC 972  
QY 949 GACACCTGTGTGAGAAAGCGCAACCCCGACTGCAAGACCATCTGCGCTCTCGGC 1008  
Db 973 GAGACCTGTGTGAGAAAGCGCAACCCCGACTGCAAGACCATCTGAAAGCTCTCGGC 1032  
QY 1009 CCCGGCGCCAGCCTGGAAGAGATGAGACCGCTGCGCAGGGCGTGGCGGCCCGCAGCCAC 1068  
Db 1033 CCCGGCGCCAGCCTTGAAGAGATGAGACCGCTGCGCAGGGCGTGGCGGCCCGCAGCCAC 1092  
QY 1069 AAGGCCCGCGTGTGGCCGAGGCGATGAGCCAGGCCAACCA-----CCAGCGTGATGATG 1122

DB 1093 AAGGCCCGCTGTGCGCCGAGGCGATGAGCCAGGTGACGAACCCGCGCATCATGATG 1152  
QY 1123 CAGAAAGCAACTTCAAGGGCCCCCGCGCATCGTCAAGTGTCTTCAACTGCGGCAAGGAG 1182  
DB 1153 CAGCGCGGCAACTTCCGCAACCAAGCGGAAGACCGTCAAGTGTCTTCAACTGCGGCAAGGAG 1212  
QY 1183 GGCACATGCGCCCGCACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGGCAAG 1242  
DB 1213 GGCCACACCGCCAGAGAACTGCGCGCGCCCCCGCAAGAGGGCTGTGCGCGCGCGC 1272  
QY 1243 GAGGCGCCACCAAGATGAGAGACTGTGACCGAGCGCCAGGCCCACTTCTGTGGCAAGATCTGG 1302  
DB 1273 GAAGGACACCAAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTAAAGGAAGATCTGG 1332  
QY 1303 CCGAGCCACAAAGGCGCGCCCCCGCAACTTCTGTGAGAGCGCGCCGAGCCCAAGCCGCC 1362  
DB 1333 CCTTCTTACAAGGGAAGGCGCAGGGAATTTCTTCAAGAGCAGACCAAGCCCAAGCCCCA 1392  
QY 1363 CCGCGCCGAGAGCTTCCGCTTCCGAGAG-----ACCAACCCCGCGCGCAAGCAGAGAG--- 1413  
DB 1393 CCAGAAAGAGAGCTTCAAGTTTGGGAGGAGAAACCAACTCCCTCTCAAGAGCAGAGAGCCG 1452  
QY 1414 --AGCAAGGACCGCGAGACCCCTGACCAAGCTGTAAGAGCCTGTTCCGCAACGAGCCCTG 1470  
DB 1453 ATAGACAAGGAAGTGTATCTTTAACTTCCCTCAGATCACTTTTGGCAACGAGCCCTCG 1512  
QY 1471 AGCCAGTAA 1479  
DB 1513 TCACAGTAA 1521

## RESULT 10

US-09-475-515-6  
; Sequence 6, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIT, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 4319  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-gag-polymerase  
US-09-475-515-6

Query Match 75.4%; Score 1114.6; DB 4; Length 4319;  
Best Local Similarity 85.8%; Pred. No. 6.6e-164;  
Matches 1295; Conservative 0; Mismatches 184; Indels 30; Gaps 4;

QY 1 ATGGGCGCGCGCGCAAGCATCTGCGCGCGCGGCAAGCTGAGCGCTGGGAGCGCATCCGC 60  
DB 7 ATGGGCGCGCGCGCGCAAGCATCTGCGCGCGCGGCGGCGAGCTGACCAAGTGGAGGAAGATCCGC 66  
QY 61 CTGGCGCGCGCGCGGCAAGAGTGTATCATGATGAAGCACTGTGTGTGGCCAGCGCGGAG 120  
DB 67 CTGGCGCGCGCGCGGCAAGAGAGTATCAAGCTGAAGCACTGTGTGTGGCCAGCGCGGAG 126

QY 121 CTGGAGAAAGTTGCGCCCTGAACCCCGGCGCTGTGAGAGACCAGCGAGGGCTGCAAGCAGATC 180  
DB 127 CTGGAGCGCTTTCGCGGTGAACCCCGGCGCTGTGAGAGACCAGCGAGGGCTGCGGCCAGATC 186  
QY 181 ATCCGCCAGCTGCAACCCCGCGCTGCAGAGCCGGCAGCGAGAGCTGAAGAGCTGTTCAC 240  
DB 187 CTGGGCCAGCTGCAACCCCGCGCTGCAGAGCCGGCAGCGAGAGCTGCGCAGCTGTACAA 246  
QY 241 ACCGTGGCCACCTGTACTGTGCTGCAGCAGAGAAATCGAGGTCCGCGACACCAAGAGGCC 300  
DB 247 ACCGTGGCCACCTGTACTGTGCTGCAGCAGCGCATCGACGTCAAGAGACCAAGAGAGGCC 306  
QY 301 CTGGACAAGATCGAGAGAGAGAGCAACAAGTCCAGCAGAGATCCAGCAGCGCGAGGCC 360  
DB 307 CTGGAGAAAGATCGAGAGAGAGAGCAACAAGTCCAGAGAGAGGCCAGCAGCGCGCGCC 366  
QY 361 GCGG-----ACAAGGCGAAGGTGAGGCCAGAACTAACCCCATCGTGCAGAACTG 408  
DB 367 GCGGCGCGCACCGGCAACAGCAGCCAGGTGAGCCAGAACTAACCCCATCGTGCAGAACTG 426  
QY 409 CAGGGCCAGATGTGTCAACCAAGGCCATCAGCCCGCGCACCTGTAAAGCTGTGGTGAAGTG 468  
DB 427 CAGGGCCAGATGTGTCAACCAAGGCCATCAGCCCGCGCACCTGTAAAGCTGTGGTGAAGTG 486  
QY 469 ATCGAGGAAAGGCTTCAAGCCCGCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGC 528  
DB 487 GTGAGGAGAAAGGCTTCAAGCCCGCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGGC 546  
QY 529 GCCACCCCGCGAGACCTGAACAGATGTTGAACACCGTGGGCGGCACCAAGCGCCCATG 588  
DB 547 GCCACCCCGCGAGACCTGAACAGATGTTGAACACCGTGGGCGGCACCAAGCGCCCATG 606  
QY 589 CAGATGCTGAAGAGACCAATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCACCCCGTG 648  
DB 607 CAGATGCTGAAGAGAGACCAATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCACCCCGTG 666  
QY 649 CAGCGCGCGCGCATCGCCCCCGCGCAGATGCGCGAGGCCCGCGCGCAGCGACATCGCGGC 708  
DB 667 CAGCGCGCGCGCATCGCCCCCGCGCAGATGCGCGAGGCCCGCGCGCAGCGACATCGCGGC 726  
QY 709 ACCACCAAGACCTTCAGAGAGAGATCGCTGATGACCAAGACCCCGCATCCCGTG 768  
DB 727 ACCACCAAGACCTTCAGAGAGAGATCGCTGATGACCAAGACCCCGCATCCCGTG 786  
QY 769 GCGGACATCTCAAGCGGTGATCATCTGCGCTGAACAAGATCGTGGATGTACAGC 828  
DB 787 GCGGACATCTCAAGCGGTGATCATCTGCGCTGAACAAGATCGTGGATGTACAGC 846  
QY 829 CCGGTGAGCATCTGAGACATCAAGAGGCGCGCAAGAGCCCTTCGCGCATCGTGAC 888  
DB 847 CCGGACGATCTCTGAGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 906  
QY 889 CGCTTCTTCAAGACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948  
DB 907 CGCTTCTTCAAGACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966  
QY 949 GACACCTGTGTGAGAGAGCGCAACCCCGCATCTGCAAGACATCTGCGCGCTCTCGC 1008  
DB 967 GAGACCTGTGTGAGAGAGCGCAACCCCGCATCTGCAAGACATCTGAGGCTCTCGC 1026  
QY 1009 CCGGCGCGCGCGCTGAGAGAGATGATGACCGCTGCGAGGCGTGGCGCGCGCGCGCGCG 1068  
DB 1027 CCGGCGCGCGCGCTGAGAGAGATGATGACCGCTGCGAGGCGTGGCGCGCGCGCGCGCG 1086  
QY 1069 AAGCGCGCGTGTGCGCGAGCGATGAGCGCAAGCA-----CCAGCGTATGATG 1122  
DB 1087 AAGCGCGCGTGTGCGCGAGCGATGAGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1146  
QY 1123 CAGAAAGCACTTCAAGGCGCGCGCGCGCGCATGCTCAAGTGTCTCAACTGCGCGCAAGAG 1182  
DB 1147 CAGCGCGCGCACTTCGCAACCAAGCGGAAGACCGTCAAGTGTCTCAACTGCGCGCAAGAG 1206  
QY 1183 GGCCACATCGCGCGCAACTGCGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGCAAG 1242

```
Db 1207 GGGCCACACCGCCAGAACTGCGCGCGCCCCCGCAGAGAGGCTGCTGGCGCTGGCGCGC 1266
QY 1243 GAGGGCCACAGATGAAGAGACTGACCGAGCGCCAGCCAACTTCCTGGCAAGATCTGG 1302
Db 1267 GAAGGACACCAATAAGAAATGACTGAGAGACAGGCTAATTTTAAAGAAATCTGG 1326
QY 1303 CCCAGCCACAAGGCGCGCCCGGCAACTTCCTGAGAGCGCGCCGAGCCCAAGCCCGCC 1362
Db 1327 CCTTCCTACAAGGGAAGGCGCAGGGAATTTCTTCAGAGCAGACCGCAAGCCCA 1386
QY 1363 CCGCGCAGAGCTTCGCTTCAGAGG-----ACCACCCCCGGCGCAGAGAGG--- 1413
Db 1387 CCAGAAAGAGCTTCAGGTTTGGGAGAGAGAAACAACCTCCCTCAGAGCAGAGCGCG 1446
QY 1414 ---AGCAAGGACCGCGAGACCTGACCAAGCTGAAGAGCTGTCGGAACGACCCCTG 1470
Db 1447 ATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTGGCAACGACCCCTCG 1506
QY 1471 AGCCAGTAA 1479
Db 1507 TCACAGTAA 1515
```

## RESULT 11

```
US-09-475-515-9
; Sequence 9, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1268
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag
; OTHER INFORMATION: common region
US-09-475-515-9
```

Query Match 69.3%; Score 1025.2; DB 4; Length 1268;  
Best Local Similarity 89.6%; Pred. No. 4.2e-150;  
Matches 1131; Conservative 0; Mismatches 113; Indels 18; Gaps 2;

```
QY 1 ATGGGGCGCGCGCCAGCATCCTGCGCGCGCGGCAAGCTGGACGCTGGAGCGCATCCGC 60
Db 7 ATGGGGCGCGCGCCAGCGTGTGAGCGCGCGCGGCTGGACAAGTGGGAGAAATCCGC 66
QY 61 CTGGCGCGCGCGCGCAAGAGTGTCTATCATGATGAAGCACTGTGTGGGCCAGCGCCGAG 120
Db 67 CTGGCGCGCGCGCGCAAGAGATGACAGCTGAAGCACATCGTGTGGGCCAGCGCCGAG 126
QY 121 CTGGAAGAAGTTCGCGCTGAACCGCGCGCTGTGAGACCAAGCGAGGCTGCAAGCAGATC 180
Db 127 CTGGAAGCGCTTCGCGGTGAACCGCGCGCTGTGAGACCAAGCGAGGCTGCGCGCAGATC 186
QY 181 ATCCGCCAGCTGCACCGCGCGCTGACAGCCGGCAGCGAGGAGCTGAAGAGCTGTTCAAC 240
Db 187 CTGGGCCAGCTGCAGCGCCAGCTGACAGCCGGCAGCGAGGAGGCTGCGCAAGCTGTAAAC 246
```

```
QY 241 ACCGTGACCACTCTGTACTGCGTGCACAGAGATTCGAGGTCCGCGACACCAAGAGGCC 300
Db 247 ACCGTGACCACTCTGTACTGCGTGCACAGCGCATCGACGTCAAGAGACCAAGAGAGGCC 306
QY 301 CTGGAAGAAGTTCGAGAGAGAGAGCAACAAGTCCAGAGAGATCCAGAGCGCGAGGCC 360
Db 307 CTGGAAGAAGTTCGAGAGAGAGAGCAACAAGTCCAGAGAGAGAGAGAGAGAGAGAGGCC 366
QY 361 GCGG-----ACAAAGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAAACTG 408
Db 367 GCGCGCGCAGCGCAACAGAGCCAGGTGAGCCAGAACTACCCCATCTGTGCAAACTG 426
QY 409 CAGGGCCAGATGCTGCACAGGCGCATCAGCCCGCGCACCTGAAAGCGCTGGTGAAGGTG 468
Db 427 CAGGGCCAGATGCTGCACAGGCGCATCAGCCCGCGCACCTGAAAGCGCTGGTGAAGGTG 486
QY 469 ATCGAGAGAGAGCGCTTCAGCGCCGAGGTGATCCCATATGTTCAACCGCCCTGAGCGAGGC 528
Db 487 GTGGAAGAGAGAGCGCTTCAGCGCCGAGGTGATCCCATATGTTCAACCGCCCTGAGCGAGGC 546
QY 529 GCCACCCCCCAGACCTGAACAACGATGTTGAACAACCGTGGCGGCCACCAAGCGCCCATG 588
Db 547 GCCACCCCCCAGACCTGAACAACGATGTTGAACAACCGTGGCGGCCACCAAGCGCCCATG 606
QY 589 CAGATGCTGAAGAGACACCATCAACGAGAGCGCGCGCGAGTGGGACCGCGTGCACCCCGTG 648
Db 607 CAGATGCTGAAGAGAGACCATCAACGAGAGCGCGCGCGAGTGGGACCGCGTGCACCCCGTG 666
QY 649 CACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708
Db 667 CACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
QY 709 ACCACCAAGCACTCTGCAAGAGAGATCGCTGATGACCAAGCAACCCCCCATCCCGTG 768
Db 727 ACCACCAAGCACTCTGCAAGAGAGATCGCTGATGACCAAGCAACCCCCCATCCCGTG 786
QY 769 GCGGACATCTAAGCGGTGATCATCTGGGCTGAAACAAGATCGTGGGATGTACAGC 828
Db 787 GCGGAGATCTAAGCGGTGATCATCTGGGCTGAAACAAGATCGTGGGATGTACAGC 846
QY 829 CCGGTGAGCATCTTGACATCAAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 888
Db 847 CCGACCAAGCATCTTGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 906
QY 889 CGCTTCTTCAAGACCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948
Db 907 CGCTTCTAAGACCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966
QY 949 GACACCTGTGTGTGCAAGAGCGCAACCCCGCATGCAAGACCATCTGCGCGCTCTGCGC 1008
Db 967 GAGACCTGTGTGTGCAAGAGCGCAACCCCGCATGCAAGACCATCTGAAAGCTCTCGGC 1026
QY 1009 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1068
Db 1027 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1086
QY 1069 AAGGCGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
Db 1087 AAGGCGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1146
QY 1123 CAGAAAGCAACTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182
Db 1147 CAGCGCGCAACTTCCGCAACGAGCGGAAGACGTCAGTGTCTTCAACTGCGGCAAGAG 1206
QY 1183 GCGCACATCGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCAAGAGGCGTGTGAGTGGCGCAAG 1242
Db 1207 GCGCACACCGCGCAAGAACTGCGCGCGCGCGCGCGCGCGCGCGCAAGAGGCGTGTGCGCGCGCG 1266
QY 1243 GA 1244
Db 1267 GA 1268
```

```
RESULT 12
US-09-952-060-27
; Sequence 27, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emimi, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Silver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR FILING DATE: 2001-09-14
; PRIOR FILING DATE: 2001-09-07
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Codon optimized DNA encoding human HIV-1 gag
US-09-952-060-27

Query Match      66.0%; Score 976.6; DB 4; Length 1521;
Best Local Similarity 80.2%; Pred. No. 1.4e-142;
Matches 1205; Conservative 0; Mismatches 274; Indels 24; Gaps 4;

QY      1 ATGGGCGCCGCGCGCAGCATCTCGCGCGCGGCAAGCTGACGCGCTGGAGCGCATCCGC 60
DB      1 ATGGGTGCTAGGGCTTCTGTGCTGTCTGTGTGTGAGCTGACAAAGTGGAGAAATCAGG 60
QY      61 CTGCGCCCCGCGCGCAAGAGTGTATCATGATGAAGCACTGTGTGGGCCAGCCGCGAG 120
DB      61 CTGAGGCGCTGTGTGCAAGAAAGATGAACAAGTAAAGCAATTGTGGGCTTCAAGGAG 120
QY      121 CTGAGAGAAGTTCGCCCTGAACCCCGGCTGTGAGAGCCAGCGAGGCTGCAAGCAGATC 180
DB      121 CTGAGAGAGTTTGTCTGTGAACCTGTGGCTGTGAGAGCTTGAAGGGGTGACAGCAGATC 180
QY      181 ATCCGCGAGCTGCAACCCGCGCTGCGAGACCGCGACGAGAGCTGAAGAGCTGTTCAC 240
DB      181 CTGGGCGAGCTCCAGCCCTCCCTGCAACAAGGCTCTGAGAGCTGAGAGTCCCTGTACAAC 240
QY      241 ACCGTGGCACCCTGTACTGTGCTGCGACGAGAAGATCGAGTCCGCGACACCAAGAGGCC 300
DB      241 ACAGTGGCTAACCTGTACTGTGTGACCAAGAAGATGTATGTGAAGACACCAAGAGGCC 300
QY      301 CTGACAAGATCGAGAGAGAGCAAGTCAAGTCCAGCAAGAATCCAGCAGGCGAGG-- 358
DB      301 CTGGAAGAAGATTGAGAGAGAGCAAGTCAAGTCAAGAAGAGGCCACAGGCTGTGCT 360
QY      359 -----CGCGCAAGAGGCAAGGTGAGCCAGAACTACCCCATCTGTGCAAGAACTGAGGGC 414
DB      361 GGCACAGGCAACTCCAGCGAGGTGTCCAGAACTACCCCATTTGTGCAAGAACTCAAGGGC 420
QY      415 CAGATGTGTACCAAGGACATCAGCCCCCGACACCTGAAGCGCTGGGTGAAGTGAATCAG 474
DB      421 CAGATGTGTACCAAGGACATCTCCCCCGGACCCCTGAATGCTTGGGTGAAGTGTGAG 480
```

```
QY      475 GAGAAAGCCTTCAACCCCGAGGTATCCCATGTTCACCGCCCTGAGCGAGGGCGCCACC 534
DB      481 GAGAAAGCCTTCTCCCTGAGGTATCCCATGTTCCTGCGCCCTGTGAGGGTGCACACC 540
QY      535 CCCGAGACCTGAACACGATTTGAACACCGTGGGCGGCCACAGCGCCCATGACAGATG 594
DB      541 CCCGAGACCTGAACACCATGTGTGAACACAGTGGGGGCCCATGAGGCTGCATGACAGATG 600
QY      595 CTGAAGGACCATCAACGAGAGAGCGCCGAGTGGGACCCGCTGACCCCGTGCAGGCC 654
DB      601 CTGAAGGACCATCAATGAGAGAGGCTGTGAGTGGGACAGGCTGCATCTGTGCAGCGT 660
QY      655 GGGCCCATCGCCCCCGCGCAGATGCGGAGCCCGCGGACGACATCGCCGACACACC 714
DB      661 GGGCCCATTTGCCCCCGCGCAGATGAGGAGGCCAGGGGCTGTGACATTGTGCGACACC 720
QY      715 AGCACCTTGAGAGAGAGATGCGCTTGATGACCAAGCAACCCCATCCCGTGGGCGAC 774
DB      721 TCCACCTTCAGAGAGAGATGTGGCTGATGACCAACACCCCATCCCTGTGGGGAA 780
QY      775 ATCTAAGCGGTGATCATCTGGGCTGGAACAAGATCGTGGAGTGTACAGCCCGTG 834
DB      781 ATCTAAGAGGTGATCATCTGGGCTGGAACAAGATGTGAGATGTACTCCCCACC 840
QY      835 AGCATCTTGACATCAAGAGAGGCCCCAAGAGCCCTTCGCGACTACGTGACCGCTTC 894
DB      841 TCCATCTTGACATCAGAGAGGCCCCAAGAGCCCTTCAGGGAATATGTGACAGGTTCC 900
QY      895 TTCAAGACCTTGGCGCGCGAGCAGAGACACCCAGAGGTGAAGAACTGATGACGACACC 954
DB      901 TACAAGACCTTGAAGGCTGAGCAGGCTCCAGAGAGTGAAGAACTGATGACAGAGACC 960
QY      955 CTGCTGTGCAAGACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCGGC 1014
DB      961 CTGCTGTGCAAGATGCAACCTGACTGCAAGACCATCTGAAAGCCCTGCGCCCTGCT 1020
QY      1015 GCCAGCCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1074
DB      1021 GCCAGCCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
QY      1075 CGCGTCTGCGCGAGGCGATGAGCCAGGCCACA-----CCAGCGTATGATGACAGAAG 1128
DB      1081 AGGGTCTGTGAGGCGCATGTCTCCAGGTGACCAACTCCGCCACCATCATGATGACAGAG 1140
QY      1129 AGCAACTTCAAGGGCCCCCGCGCATGTCAAGTCTTCAACTGCGGCAAGAGGGCCAC 1188
DB      1141 GCGCACTTCAAGAACAGAGAGAGACAGTGAAGTCTTCAACTGTGCAAGGTGGGCCAC 1200
QY      1189 ATCGCCCGCACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGAGAGGGC 1248
DB      1201 ATTGCCAAGACTGTAGGGCCCCCAGGAAGAGGGCTGTGGAAGTGTGCAAGAGAGGGC 1260
QY      1249 CACCAATGAAGACTGCACCGAGCGCCCACTTCTGTGGGCAAGATCTGGCCAGC 1308
DB      1261 CACCAATGAAGACTGCATGAAGAGGCAAGGCACTTCTGTGGGCAAAATCTGGCCCTCC 1320
QY      1309 CACAAGGGCCCCCGGCACTTCTGCAAGAGCCCGCCAGCCCAAGCCCGCCCGCCGCC 1368
DB      1321 CACAAGGGCAGGCTTGCAACTTCTCAGTTCAGGCTGAGAGCCCAAGGCCCTCCGAG 1380
QY      1369 GAGAGCTTCCGCTT-----GAGGAGACCAACCCCGGCGCAAGAGAG-----AGC 1416
DB      1381 GAGTCTTCAAGTTTGGGAGAGAGAGACCAACCCCGGCAAGAGAGGCCATTGAC 1440
QY      1417 AAGGACCGCAGACCTTGACCAAGCTGAAAGAGCTGTTCCGCAAGACCCCTGAGCCAG 1476
DB      1441 AAGGAGCTGTACCCCTTGCGCTTCCCTGAGGTTCCCTGTTGGCAAGACCCCTTCCAG 1500
QY      1477 TAA 1479
DB      1501 TAA 1503
```

```

RESULT 13
US-09-974-702-1
; Sequence 1, Application US/09974702
; Patent No. 6696291
;
GENERAL INFORMATION:
; APPLICANT: Shiver, John W.
; Davies, Mary Ellen
; Freed, Daniel C.
; Liu, Margaret A.
; Perry, Helen C.
;
TITLE OF INVENTION: Synthetic HIV Gag Genes
;
NUMBER OF SEQUENCES: 1
;
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0900
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/974,702
; FILING DATE: 09-Oct-2001
; CLASSIFICATION: <Unknown>
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/017,981
; FILING DATE: 03-FEB-1998
; APPLICATION NUMBER: US60/037,854
; FILING DATE: 07-FEB-1997
; APPLICATION NUMBER: GB9705040.5
; FILING DATE: 12-MAR-1997
;
ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19730
;
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-3905
; TELEFAX: 732-594-4720
; TELEX: <Unknown>
;
INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
;
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-974-702-1
;
Query Match      66.0%; Score 976.6; DB 4; Length 1532;
Best Local Similarity 80.2%; Pred. No. 1.4e-142;
Matches 1205; Conservative 0; Mismatches 274; Indels 24; Gaps 4;
;
QY      1 ATGGGCGCCCGCCGCAAGATCCTTCGCGCGCGGCGGCAAGCTGGACGCGCTGGAGCGCATCCGC 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10 ATGGGTGCTAGGGCTTCTGTGCTGTCTGTGTGCTGTGAGCTGGACAAGTGGAGAGATCAGG 69
;
QY      61 CTGCGCCCCCGCGGCAAGAGTGTATCATGATGTAAGCACCTGGTGTGGGCCAGCCGCGAG 120
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      70 CTGAGGCCCTGTGGCAAGAGAGTACAGACTAAAGACATTGTGTGGCCCTCCAGGGAG 129
;
QY      121 CTGAGAGAAGTTGCGCCTGAACCCCGGCTGTCTGGAGACCAGCGAGGGCTGCAAGCAGATC 180
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      130 CTGAGAGAAGTTTGTCTGTGAACCCCTGGCTGTCTGAGAGACCTCTGAGGGGCTGCAAGCAGATC 189
;
      181 ATCCGCCAGCTGCACCCCGCCCTGTCAGACCGGCGAGCGAGGAGCTGAAGAAGCCTTTCAAC 240

```

D	b	190	CTGGGCGAGCTCCAGCCCTCCCTGCAACAGGCTCTGAGAGCTGAGGTCCCTGTACAAC	249
O	y	241	ACCGTGGCCACCCTGTACTGCGTGCAAGAAAGATCGAGGTCCGCGCACCAAGAGAGCC	300
D	b	250	ACAGTGGCTACCTGTACTGTGTGCAACGAAAGATTGATGTGAAGGACACCAAGAGAGCC	309
O	y	301	CTGGAACAAGATCGAGAGAGAGCAAGCAAGTGCAGACGAAAGATCCAGAGGCCGAGG--	358
D	b	310	CTGGAGAAGATTGAGAGAGAGCAACAAGTCCAAGAAAGAGGCCACAGAGCTGTGCT	369
O	y	359	----CCGCCGACAAGGCGAAGGTGAGCCAGAACTAACCCCATCTGTGAGAACCTGCAAGGC	414
D	b	370	GGCAAGGCAACTCCAGCCAGGTGTCCAGAACTAACCCCATTTGTGCAAGAACTCCAGGGC	429
O	y	415	CAGATGTTGACACGAGCCATCAGCCCCGCAACCTGAACGCGCTGGGTGAAGTATCGAG	474
D	b	430	CAGATGTTGACACGAGCCATCTCCCCGCAACCTGAATGCTGGGTGAAGTGTGGAG	489
O	y	475	GAGAAAGCCTTCAGCCCCGAGGTGATCCCATGTTACCGCCCTGAGCGAGGCGCCACC	534
D	b	490	GAGAAAGCCTTCTCCCTGAGGTGATCCCATGTTCTCTGCCCTGTCTGAGGTTGCCACC	549
O	y	535	CCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCACAGGCGCCATGAGATG	594
D	b	550	CCCCAGGACCTGAACACCATGTGAAACACAGTGGGGGGCCATCAGGCTGCCATGAGATG	609
O	y	595	CTGAAGGACACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCAACCCGTGACGCC	654
D	b	610	CTGAAGGAGACCATCAATGAGAGGCTGTGAGTGGGACAGGCTGTCATCCTGTGACGCT	669
O	y	655	GGCCCCATCGCCCCCGGCGAGATGCGCGAGGCCCGCGAGCGACATCGCCGCGCACACC	714
D	b	670	GGCCCCATTCGCCCGCGCGAGATGAGGAGGCCAGGGGCTTGACATTTGCTGGCACACC	729
O	y	715	AGCACCTCTGAGAGCAGATCGCCTGATGACCAACACCCCATCCCGTGGGCGAC	774
D	b	730	TCCACCTCCAGAGCAGATTGGCTGATGACCAACACCCCATCCCTGTGGGGAA	789
O	y	775	ATCTACAAGCGGTGATCATCTTGCGCCTGAACAAGATCGTGCGATGTACACCCCGTG	834
D	b	790	ATCTACAAGGTTGATCATCTTGCGCCTGAACAAGATTGTGAGATGTACTCCCCACC	849
O	y	835	AGCATCCTGACATCAAGCAGGCGCCCAAGAGCCCTTCGCGCACTACGTGACCGCTTC	894
D	b	850	TCCATCCTGACATCAGGAGGCGCCCAAGAGCCCTTCAGGGACTATGTGACAGGTTTC	909
O	y	895	TTCAGAAGCCTGCGCGCGAGCAGACACCCAGAGGTGAAGACTGATGACCGACACC	954
D	b	910	TACAAGACCTGAGGGCTGAGCAGGCTCCCAAGAGGTGAAGACTGATGACAGAGACC	969
O	y	955	CTGCTGTGAGAAAGCCAAACCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCGGC	1014
D	b	970	CTGCTGTGAGAAATGCCAACTGACTGCAAGACCATCTGAAGGCCCTGGGCGCTGCT	1029
O	y	1015	GCCAGCCTGAGAGATGATGACCGCCTGCAAGGCGTGGGCGGCCCCAGCCCAAGGCC	1074
D	b	1030	GCCACCTGAGAGATGATGACAGCCTGCCAGGGGGTGGGGGCCCTGTGTCAAGGCC	1089
O	y	1075	CGCGTGTGCGCGAGCGGATGAGCCAGGCCAACCA-----CCAGCTGATGATGAGAG	1128
D	b	1090	AGGGTGTGCTGAGGCGCATGTCCAGGTGACCAACTCCGCCACATCATGATGACAGAG	1149
O	y	1129	AGCAACTTCAAGGGCCCCCGCGCATGTCAAGTGTCTTCAACTGCGGCAAGAGGGCCAC	1188
D	b	1150	GGCACACTTCAGGAACAGAGGAAGACAGTGAAGTCTTCAACTGTGGCAAGGTGGCCAC	1209
O	y	1189	ATCGCCGCAACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGAGAGGGC	1248
D	b	1210	ATTGCCAAGAACTGTAGGGCCCCCAGGAAGAGGGCTGTGGAAGTGTGCAAGAGGGC	1265
O	y	1249	CACCATGTAAGGACTGACCGAGCGCCAGGCCAACTTCTGGGCAAGATGTGGCCAGC	1308



QY 1369 GAGAGCTTCCGCTT-----CGAGGAGACCACCCCGGGCCAAGAAGCAGGAG-----AGC 1416  
||| ||| |  
Db 1390 GAGTCCTTCAGGTTTGGGGAGGAGAAAGACCACCCCCAGCCAGAAGCAGGAGCCATTGAC 1449  
  
QY 1417 AAGGACCGCGAGACCCTGAACCAAGCCTGTTCGGCAACGACCCCCTGAGCCAG 1476  
||||| ||| |  
Db 1450 AAGGAGCTGTACCCCTGGCTCCCTGAGGTCCCTGTTGGCAACGACCCCTCTCCAG 1509  
  
QY 1477 TAA 1479  
|||  
Db 1510 TAA 1512

```

RESULT 15
US-09-952-060-25
; Sequence 25, Application US/09952060
; Patent No. 6733993
; GENERAL INFORMATION:
; APPLICANT: Emin, Emilio A.
; APPLICANT: Youll, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shaver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Castimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 37474
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DNA encoding pMRKads HIV-1 gag, coding
US-09-952-060-25

```

	Query Match	66.0%;	Score 976.6;	DB 4;	Length 37474;	
	Best Local Similarity	80.2%;	Pred. No. 1.4e-142;			
	Matches 1205;	Conservative	0;	Mismatches 274;	Indels 24;	Gaps 4;
Qy	1 ATGGCGCCCGCGCCAGCATCTGCGCGCGCAACTGACGCCTGGAGCGCATCCGC	60				
Db	1272 ATGGTGCTAAGGGCTTCTGTGCTGTCTGTGTGTGAGCTGACAACTGGAGAAGATCAGG	1331				
Qy	61 CTGCGCCCCCGCGCGCMAAGTGTCTACATGATGAAGCACCTGTGTGGCCACGCCGAG	120				
Db	1332 CTGAGCCTGTGTGGCMAAGAAAGTACAAGCTAAAGCACATTGTGTGGCCTTCAGGGAG	1391				
Qy	121 CTGAGAAGTTGCCCCTGAACCCCGGCTGTGTGAGACCAGCGAGGGCTGCMAAGCATC	180				
Db	1392 CTGAGAAGTTTGTGTGAACCTGGCCTGTGTGAGACCTCTGAAGGGGTGACGCGAGATC	1451				
Qy	181 ATCCGCGAAGCTGCACCCCGCCCTGCAGACCAGCGACGAGAGCTGAAGAGCTGTTCAC	240				
Db	1452 CTGGGCGAGCTCCAGCCCTCCCTGCMAAACAGAGCTTGTAGAGAGTGAAGTCCCTGTACAAC	1511				
Qy	241 ACCGTGCGCACCTGTACTGCGGTGCACGAGAAGATCGAGTCCGCGACACCAAGAGAGCC	300				

Db	1512	ACAGTGGCTACCCCTGTACTGTGTGCACCAGAAAGATTGATGTGAAGGACACCAAGAGGCC	1571
OY	301	CTGGACAAGATCGAGGAGGACAGAACAAAGTGCACAGAAAGATCCAGCAGGCCGAGG--	358
Db	1572	CTGGAGAAGATTGAGGAGGACAGAACAAAGTCCAAAGAAAGGCCCAAGCAGCTGCTGCT	1631
OY	359	----CCGCCGCAAGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACTCTCAGAGGC	414
Db	1632	GGCACAGGCGCACTCCAGCCAGGTGTCCAGAACTACCCCATTTGTGCAGAACTCCAGGGC	1691
OY	415	CAGATGTGCACCAAGGCCATCAGCCCCCGCACCTTGAACGCTTGGGTGAAGGTGATCGAG	474
Db	1692	CAGATGTGCACCAAGGCCATCTCCCCCGCACCTTGAATGCTTGGGTGAAGGTGTGAG	1751
OY	475	GAGAAAGGCTTTCAGCCCCCGAGGTGATCCCCATGTTCAACGCTTGAAGGAGGCCACC	534
Db	1752	GAGAAAGGCTTTCCTCCCTGAGGTGATCCCCATGTTCTCTGCTGTGAGGAGTGCACC	1811
OY	535	CCCCAGGACCTGAACACGATGTGAACACCGTGGGGGCCCAACAGGCCGCCATGCAGATG	594
Db	1812	CCCCAGGACCTGAACACCGATGTGAACACAGTGGGGGCCCATCAGGCTGCCATGCAGATG	1871
OY	595	CTGAAGGACACCATCAACGAGAGAGGCCGCCGAGTGGGACCGCGTGCACCCCGTGCACGCC	654
Db	1872	CTGAAGGAGACCATCAATGAGAGAGGCTGCTGAGTGGGACAGAGCTGCATCTCTGTCAAGCT	1931
OY	655	GCCCCCATCGCCCCCGGCGCAGATGCGCGAGCCCCCGGCGAGCGACATCGCCGCAACACC	714
Db	1932	GCCCCCATTTGCCCCCGGCGCAGATGAGGAGGCCCAAGGGCTCTGAACATTTGCTGCAACACC	1991
OY	715	AGCACCTGTGAGGAGCAGATGCGCTTGATGACCAAGCAACCCCCCATCCCGTGGCGAC	774
Db	1992	TCCACCTCTCAAGAGCAGATTGGCTGATGACCAACACCCCCCATCCCTGTGGGGAA	2051
OY	775	ATCTACAAGCGGTGGATCATCTTGGGCTTGAACAAGATCGTGGCGATGTACAGCCCCGTG	834
Db	2052	ATCTACAAGAGTGGATCATCTTGGGCTTGAACAAGATTGTGAGGATGTACTCCCCCACC	2111
OY	835	AGCATCTTGACATCAAGCAGAGGCCCCCAAGAGCCCTTCCCGCACTACGTGACCGCTTC	894
Db	2112	TCCATCTTGACATCAGGCGAGGGCCCCAAGAGCCCTTCAGGGACTATGTGACAGGTTTC	2171
OY	895	TTCAAGACCCCTGCGCGCGCGAGCAGACCAACCAAGAGGTGAAGAACTGATGACCGACACC	954
Db	2172	TACAAAGACCTCTGAGGGCTGAGCAGGCCCTCCCAAGAGGTGAAGAACTGATGACAGAGCC	2231
OY	955	CTGCTGTGTCAGAACGCCAACCCCGCACTGCAAGACCATCTCTGCGCGCTCTCGGCCCGGC	1014
Db	2232	CTGCTGTGTCAGAAATGCCAACCCCTGACTGCAAGACCATCTCTGAAGGCCCTGSGCCCTGCT	2291
OY	1015	GCCAGCCTGAGAGAGATGATGACCCGCTGCGCAGGGCGTGGGCGGGCCCCAAGCCCAAGGCC	1074
Db	2292	GCCACCCCTGAGAGAGATGATGACAGCCTGCGCAGGGGCTGGGGGGCCCTGCTCAAGGCC	2351
OY	1075	CGCGTGTGCGCGGAGGCGGATGAGCCAGGCCAAC-----CAAGCGTATGATGCAGAG	1128
Db	2352	AGGGTGTGCTGAGAGGCCCATGTCCAGGTGACCAACTCCGCCACCATCATGATGCAGAGG	2411
OY	1129	AGCAACTTCAAGGGGCCCCCGCGGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGGCCAC	1188
Db	2412	GCGCAACTTCAGGAACCAAGAGGAAGCAAGTGAAGTGTCTCAACTGTGGCAAGGTGGGCCAC	2471
OY	1189	ATCGCCCCGCAACTGCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCAGGCAAGAGGGC	1248
Db	2472	ATTGCCAAGAACTGTAGGGCCCCCAAGGAAGGGCTGCTGGAAGTGTGGAAGAGAGGGC	2531
OY	1249	CACCAAGATGAAGACTGCACCGAGCGCCAAGCCAACTTCTTGGGCAAGATCTGGCCAGC	1308
Db	2532	CACCAAGATGAAGACTGCATATGAGAGGCAAGCCAACTTCTTGGGCAAAATCTGGCCCTCC	2591
OY	1309	CACAAAGGGCGCGCCCCGCACTTCTCTGCAAGCGCGCCCCGAGCCCCACCGCCCCCGGCC	1368
Db	2592	CACAAAGGCGCGCTTGGCACTTCTCTCAAGTCCAGGCTGAGGCCACAGCCCCCTCCGAG	2651



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 21:18:31 ; Search time 897.894 Seconds  
(without alignments)  
9798.345 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 1479

Sequence: 1 atggggcgccgcgcagcat.....acgaacctgagccagtaa 1479

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 5537552 seqs, 2974263231 residues

Total number of hits satisfying chosen parameters: 11075104

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1479	100.0	1479	10	US-09-967-464-63 Sequence 63, Appl
2	1479	100.0	1479	10	US-09-899-575-3 Sequence 3, Appl
3	1477.4	99.9	3162	15	US-10-190-435-18 Sequence 18, Appl
4	1477.4	99.9	3462	15	US-10-190-435-16 Sequence 16, Appl
5	1477.4	99.9	4419	15	US-10-190-435-19 Sequence 19, Appl
6	1477.4	99.9	4419	16	US-10-190-305A-14 Sequence 14, Appl
7	1477.4	99.9	4483	15	US-10-190-435-35 Sequence 35, Appl
8	1477.4	99.9	4606	15	US-10-190-435-34 Sequence 34, Appl
9	1477.4	99.9	4615	15	US-10-190-435-36 Sequence 36, Appl
10	1477.4	99.9	4702	15	US-10-190-435-38 Sequence 38, Appl
11	1477.4	99.9	4716	15	US-10-190-435-17 Sequence 17, Appl

12	1477.4	99.9	4716	16	US-10-190-305A-13	Sequence 13, Appl
13	1475.8	99.8	2742	15	US-10-190-435-20	Sequence 20, Appl
14	1475.8	99.8	2742	16	US-10-190-305A-15	Sequence 15, Appl
15	1475.8	99.8	3930	15	US-10-190-435-9	Sequence 9, Appl
16	1475.8	99.8	3930	15	US-10-190-435-10	Sequence 10, Appl
17	1475.8	99.8	3930	15	US-10-190-435-11	Sequence 11, Appl
18	1475.8	99.8	5145	15	US-10-190-435-12	Sequence 12, Appl
19	1475.8	99.8	5145	16	US-10-190-305A-12	Sequence 59, Appl
20	1474.8	99.7	4713	15	US-10-190-435-59	Sequence 83, Appl
21	1474.8	99.7	4713	16	US-10-190-305A-83	Sequence 82, Appl
22	1473.8	99.6	5184	15	US-10-190-435-58	Sequence 58, Appl
23	1473.8	99.6	5184	16	US-10-190-305A-82	Sequence 67, Appl
24	1473.4	99.6	1479	10	US-09-967-464-67	Sequence 82, Appl
25	1472.4	99.6	2742	15	US-10-190-435-57	Sequence 57, Appl
26	1472.4	99.6	2742	16	US-10-190-305A-81	Sequence 81, Appl
27	1463	98.9	1479	10	US-09-899-575-20	Sequence 20, Appl
28	1315.2	88.9	3531	15	US-10-190-435-13	Sequence 13, Appl
29	1315.2	88.9	3537	15	US-10-190-435-14	Sequence 14, Appl
30	1315.2	88.9	3537	15	US-10-190-435-15	Sequence 15, Appl
31	1288.8	87.1	1491	10	US-09-899-575-99	Sequence 99, Appl
32	1288	87.1	1494	10	US-09-899-575-51	Sequence 51, Appl
33	1276.8	86.3	1509	10	US-09-899-575-4	Sequence 4, Appl
34	1275.2	86.2	1509	10	US-09-967-464-64	Sequence 64, Appl
35	1273.6	86.1	1509	10	US-09-967-464-68	Sequence 68, Appl
36	1260.8	85.2	1509	10	US-09-899-575-21	Sequence 21, Appl
37	1221.2	82.6	9166	17	US-10-359-120-168	Sequence 168, App
38	1206	81.5	9788	17	US-10-359-120-174	Sequence 174, App
39	1197.8	81.0	1515	17	US-10-387-336-4	Sequence 4, Appl
40	1197.8	81.0	4472	17	US-10-387-336-75	Sequence 75, Appl
41	1197.8	81.0	4608	17	US-10-387-336-76	Sequence 76, Appl
42	1197.8	81.0	4689	17	US-10-387-336-74	Sequence 74, Appl
43	1197.8	81.0	4766	17	US-10-387-336-73	Sequence 73, Appl
44	1196.2	80.9	2799	16	US-10-241-009-18	Sequence 18, Appl
45	1196.2	80.9	2799	16	US-10-190-434B-18	Sequence 18, Appl

#### ALIGNMENTS

RESULT 1  
US-09-967-464-63  
; Sequence 63, Application US/09967464  
; Publication No. US20030138453A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Hagan, Derek  
; APPLICANT: Otten, Gillis  
; APPLICANT: Donnelly, John J.  
; APPLICANT: Polo, John M.  
; APPLICANT: Barnett, Susan  
; APPLICANT: Singh, Mamohan  
; APPLICANT: Ulmer, Jeffrey  
; APPLICANT: Dubensky, Jr., Thomas W.  
; TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF HETEROLOGOUS NUCLEIC ACIDS  
; FILE REFERENCE: P16269.004  
; CURRENT APPLICATION NUMBER: US/09/967,464  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/236,105  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: 60/315,905  
; PRIOR FILING DATE: 2001-08-30  
; NUMBER OF SEQ ID NOS: 68  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 63  
; LENGTH: 1479  
; TYPE: DNA  
; ORGANISM: Human Immunodeficiency virus type 1  
US-09-967-464-63

Query Match 100.0%; Score 1479; DB 10; Length 1479;  
Best Local Similarity 100.0%; Pred. No. 4.1e-308;  
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 ATGGGCGCCGCGCCAGCATCTGCGCGGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60

```
Db 1 ATGGGCGCCCGCGCAGCATCTGCGGGCGGCAAGCTGGACGCTGGGAGCGCATCCGC 60
QY 61 CTGCGCCCGCGCGCAAGAGTGTATGATGAAGCACTGTGTGGCCAGCCGCGAG 120
Db 61 CTGCGCCCGCGCGCAAGAGTGTATGATGAAGCACTGTGTGGCCAGCCGCGAG 120
QY 121 CTGAGAAGTTGCGCCCTGAACCCCGGCTGTGAGACCAAGGAGCTGCAAGAGATC 180
Db 121 CTGAGAAGTTGCGCCCTGAACCCCGGCTGTGAGACCAAGGAGCTGCAAGAGATC 180
QY 181 ATCCGCGAGCTGACCCCGGCTGAGACCGGCAAGAGAGTGAAGAGCTGTCAAC 240
Db 181 ATCCGCGAGCTGACCCCGGCTGAGACCGGCAAGAGAGTGAAGAGCTGTCAAC 240
QY 241 ACCGTGGCCACCTGTATGCGTGCAAGAGAGTCCGCGACCAAGAGGCGC 300
Db 241 ACCGTGGCCACCTGTATGCGTGCAAGAGAGTCCGCGACCAAGAGGCGC 300
QY 301 CTGGAAGAATCGAGGAGAGCAAGCAAGTGCAGAGAGATCCAGAGGCGCGCC 360
Db 301 CTGGAAGAATCGAGGAGAGCAAGCAAGTGCAGAGAGATCCAGAGGCGCGCC 360
QY 361 GCCGCAAGGGCAAGGTGAGCCAGAACTAACCCTGCTGCAAGAACTGCAAGGCGCAGATG 420
Db 361 GCCGCAAGGGCAAGGTGAGCCAGAACTAACCCTGCTGCAAGAACTGCAAGGCGCAGATG 420
QY 421 GTGACCAAGGCCATCAAGCCCGGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480
Db 421 GTGACCAAGGCCATCAAGCCCGGCAACCTGAACGCTGGGTGAAGGTGATCGAGAGAAG 480
QY 481 GCCTTCAGCCCGGAGGTGATCCCAATGTTCACCGCCCTGAGCGAGGCGCCACCCCGCAG 540
Db 481 GCCTTCAGCCCGGAGGTGATCCCAATGTTCACCGCCCTGAGCGAGGCGCCACCCCGCAG 540
QY 541 GACCTGAACACGATGTTGAACACCGGTGGGCGGCCACAGGCGCCATGCAATGCTGAAG 600
Db 541 GACCTGAACACGATGTTGAACACCGGTGGGCGGCCACAGGCGCCATGCAATGCTGAAG 600
QY 601 GACACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCC 660
Db 601 GACACCATCAACGAGAGGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCC 660
QY 661 ATCGCCCCCGGCGAGATGCGCGAGCGCCGCGGCAAGCATCGCGGCAACCAAGCAACC 720
Db 661 ATCGCCCCCGGCGAGATGCGCGAGCGCCGCGGCAAGCATCGCGGCAACCAAGCAACC 720
QY 721 CTGCAAGAGCAGATCGCCTGATGACCAAGCAACCCCGCATCCCGTGGGCGCATCTAC 780
Db 721 CTGCAAGAGCAGATCGCCTGATGACCAAGCAACCCCGCATCCCGTGGGCGCATCTAC 780
QY 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGTGAACATC 840
Db 781 AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGTGAACATC 840
QY 841 CTGACATCAAGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGAACCGCTTCTCAAG 900
Db 841 CTGACATCAAGCAGGCGCCCAAGAGCCCTTCCGCGACTACGTGAACCGCTTCTCAAG 900
QY 901 ACCCTGCGCGCGGAGCAGACCAAGAGGAGTGAAGAACTGATGACCGACCCCTGCTG 960
Db 901 ACCCTGCGCGCGGAGCAGACCAAGAGGAGTGAAGAACTGATGACCGACCCCTGCTG 960
QY 961 GTGCAAGAACGCAACCCCGACTGCAAGCAACCATCTGCGCGCTCTGGGCGCGCGCAGC 1020
Db 961 GTGCAAGAACGCAACCCCGACTGCAAGCAACCATCTGCGCGCTCTGGGCGCGCGCAGC 1020
QY 1021 CTGGAAGAGATGATGACCGCTGCGAGGGGCGTGGGCGCGCCAGCGCAAGGCGCGGTG 1080
Db 1021 CTGGAAGAGATGATGACCGCTGCGAGGGGCGTGGGCGCGCCAGCGCAAGGCGCGGTG 1080
QY 1081 CTGCGCGAGGCGATGAGCCAGGCGCAACCAAGCGTGTATGACAGAGCAACTTCAAG 1140
Db 1081 CTGCGCGAGGCGATGAGCCAGGCGCAACCAAGCGTGTATGACAGAGCAACTTCAAG 1140
```

```
Db 1081 CTGCGCGAGGCGATGAGCCAGGCGCAACCAAGCGTGTATGACAGAGCAACTTCAAG 1140
QY 1141 GGGCCCCGGCGCATGTCAGAGTCTTCAACTGCGGCAAGAGGCGCACATCGCCCGCAAC 1200
Db 1141 GGGCCCCGGCGCATGTCAGAGTCTTCAACTGCGGCAAGAGGCGCACATCGCCCGCAAC 1200
QY 1201 TGCCCGCGCCCCCGCAAGAGGAGTGTGAAGTGGCGCAAGAGGCGCACAGATGAAG 1260
Db 1201 TGCCCGCGCCCCCGCAAGAGGAGTGTGAAGTGGCGCAAGAGGCGCACAGATGAAG 1260
QY 1261 GACTGACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCTGGGCCAAGCCACAGGCGCGC 1320
Db 1261 GACTGACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCTGGGCCAAGCCACAGGCGCGC 1320
QY 1321 CCGGCAACTTCTGTCAGAGCGCGCCCGAGCGCCACCGCCCCCGCGGAGCTTCCGC 1380
Db 1321 CCGGCAACTTCTGTCAGAGCGCGCCCGAGCGCCACCGCCCCCGCGGAGCTTCCGC 1380
QY 1381 TTCGAGGAGACCAACCCCGCGCGAGAGAGCAAGAGCAAGGACCGGAGACCTGACCAAGC 1440
Db 1381 TTCGAGGAGACCAACCCCGCGCGAGAGAGCAAGAGCAAGGACCGGAGACCTGACCAAGC 1440
QY 1441 CTGAAGAGCTGTTCGCGCAAGCAACCCCTGAGCCAGTAA 1479
Db 1441 CTGAAGAGCTGTTCGCGCAAGCAACCCCTGAGCCAGTAA 1479
```

```
RESULT 2
US-09-899-575-3
; Sequence 3, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
; APPLICANT: Zur Megede, Jan
; APPLICANT: Barnett, Susan W.
; APPLICANT: Egnelbrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janse
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP01631.102
; CURRENT APPLICATION NUMBER: US/09/899,575
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475,704
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag
US-09-899-575-3

Query Match      100.0%; Score 1479; DB 10; Length 1479;
Best Local Similarity 100.0%; Pred. No. 4.1e-308;
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCCCGCGCAGCATCTGCGGGCGGCAAGCTGGACGCTGGGAGCGCATCCGC 60
Db 1 ATGGGCGCCCGCGCAGCATCTGCGGGCGGCAAGCTGGACGCTGGGAGCGCATCCGC 60
QY 61 CTGCGCCCGCGCGCAAGAGTGTATGATGAAGCACTGTGTGGCCAGCCGCGAG 120
Db 61 CTGCGCCCGCGCGCAAGAGTGTATGATGAAGCACTGTGTGGCCAGCCGCGAG 120
QY 121 CTGAGAAGTTGCGCCCTGAACCCCGGCTGTGAGACCAAGGAGCTGCAAGAGATC 180
Db 121 CTGAGAAGTTGCGCCCTGAACCCCGGCTGTGAGACCAAGGAGCTGCAAGAGATC 180
QY 181 ATCCGCGAGCTGACCCCGGCTGAGACCGGCAAGAGAGTGAAGAGCTGTCAAC 240
Db 181 ATCCGCGAGCTGACCCCGGCTGAGACCGGCAAGAGAGTGAAGAGCTGTCAAC 240
```



QY 481 GCCTTCAGCCCCGAGGTGATCCCAATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
DB 487 GCCTTCAGCCCCGAGGTGATCCCAATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 546  
QY 541 GACCTGAACACGATGTTGAACACCCGTTGGGGGGCCCAACGAGGGCGCCATGAGATGCTGAAG 600  
DB 547 GACCTGAACACGATGTTGAACACCCGTTGGGGGGCCCAACGAGGGCGCCATGAGATGCTGAAG 606  
QY 601 GACACCATCAACGAGAGAGGGCGCCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCCC 660  
DB 607 GACACCATCAACGAGAGAGGGCGCCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCCC 666  
QY 661 ATCGCCCCCGGCGAGATGCGGAGCGCGCGCGGAGCGACATCGCGGACACCAAGCAACC 720  
DB 667 ATCGCCCCCGGCGAGATGCGGAGCGCGCGCGGAGCGACATCGCGGACACCAAGCAACC 726  
QY 721 CTGCAAGAGCAGATCGCCTTGATGACCAACACCCCCCATCCCGGTGGGCGACATCTAC 780  
DB 727 CTGCAAGAGCAGATCGCCTTGATGACCAACACCCCCCATCCCGGTGGGCGACATCTAC 786  
QY 781 AAGCGGTGATCATCTGGGCTGAAACAAGATCGTGCGGATGTACAGCCCGTGAACATC 840  
DB 787 AAGCGGTGATCATCTGGGCTGAAACAAGATCGTGCGGATGTACAGCCCGTGAACATC 846  
QY 841 CTGACATCAAGCAGAGGGCCCCCAAGAGCCCTTCCGGACTACGTGAACCGTTCTTCAAG 900  
DB 847 CTGACATCAAGCAGAGGGCCCCCAAGAGCCCTTCCGGACTACGTGAACCGTTCTTCAAG 906  
QY 901 ACCCTGCGCGCGGAGCAGAGCACCAGAGGTGAAGAACTGATGACCGACACCTGTCTG 960  
DB 907 ACCCTGCGCGCGGAGCAGAGCACCAGAGGTGAAGAACTGATGACCGACACCTGTCTG 966  
QY 961 GTGCAAGACGCCAACCCCGACTGCAAGAACCATCTGCGCGCTCTGCGGCCCGCGCCAGC 1020  
DB 967 GTGCAAGACGCCAACCCCGACTGCAAGAACCATCTGCGCGCTCTGCGGCCCGCGCCAGC 1026  
QY 1021 CTGAGAGAGATGATACCGCCTGCAAGGGCGGTGGGGCGGCCCAAGCCACAGGCCCGCTG 1080  
DB 1027 CTGAGAGAGATGATACCGCCTGCAAGGGCGGTGGGGCGGCCCAAGCCACAGGCCCGCTG 1086  
QY 1081 CTGCGCGGAGCGGATGAGCCAGGCCAACACCAAGCGTATGATGACAGAAAGAACATTTCAAG 1140  
DB 1087 CTGCGCGGAGCGGATGAGCCAGGCCAACACCAAGCGTATGATGACAGAAAGAACATTTCAAG 1146  
QY 1141 GGGCCCCCGGCGCATGCTCAAGTGTCTTCACTGCGGCAAGAGGGCCACATCGCCCGCAAC 1200  
DB 1147 GGGCCCCCGGCGCATGCTCAAGTGTCTTCACTGCGGCAAGAGGGCCACATCGCCCGCAAC 1206  
QY 1201 TGCGCGCGCCCCCGCAAGAGGGGTCTGGAAGTGCGGCAAGAGGGCCACCAAGTGAAG 1260  
DB 1207 TGCGCGCGCCCCCGCAAGAGGGGTCTGGAAGTGCGGCAAGAGGGCCACCAAGTGAAG 1266  
QY 1261 GACTGCAACCGAGCGCCAGGCCAACTTCTTGGGCAAGATCTGCGCCAGCCACAGAGGCCGC 1320  
DB 1267 GACTGCAACCGAGCGCGCCAGGCCAACTTCTTGGGCAAGATCTGCGCCAGCCACAGAGGCCGC 1326  
QY 1321 CCGGCAACTTCTTGCAGAGCGCGCCGAGCCCAACCGCCCCCGCGCGAGAGCTTCCGC 1380  
DB 1327 CCGGCAACTTCTTGCAGAGCGCGCGCCGAGCCCAACCGCCCCCGCGCGAGAGCTTCCGC 1386  
QY 1381 TTGAGAGAGACCAACCCCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1387 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
QY 1441 CTGAAGAGCGCTTTCGGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
DB 1447 CTGAAGAGCGCTTTCGGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485

RESULT 4  
US-10-190-435-16  
; Sequence 16, Application US/10190435

; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 3462  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagProtiNatTmut\_C  
US-10-190-435-16

Query Match 99.9%; Score 1477.4; DB 15; Length 3462;  
Best Local Similarity 99.9%; Pred. No. 8.8e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGGGCGCGCGCGCAGATCTGCGCGCGCGCAAGCTGGAAGCGCTGGAGCGCATCCGC 60  
DB 7 ATGGGCGCGCGCGCAGATCTGCGCGCGCGCAAGCTGGAAGCGCTGGAGCGCATCCGC 66  
QY 61 CTGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGACACTGTGTGGCGCGCGAG 120  
DB 67 CTGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGACACTGTGTGGCGCGCGAG 126  
QY 121 CTGAGAAAGTTGCGCCCTGAACCCCGCGCTGTGAGAGACCAAGGGCTGCAAGCATC 180  
DB 127 CTGAGAAAGTTGCGCCCTGAACCCCGCGCTGTGAGAGACCAAGGGCTGCAAGCATC 186  
QY 181 ATCCGCGAGCTGCAACCCCGCGCTGAGAGCCGCGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 187 ATCCGCGAGCTGCAACCCCGCGCTGAGAGCCGCGAGAGAGAGAGAGAGAGAGAG 246  
QY 241 ACCGTGGCCACCTTACTGCGGTGACAGAGAGATGAGGTCCGACACCAAGAGAGCC 300  
DB 247 ACCGTGGCCACCTTACTGCGGTGACAGAGAGATGAGGTCCGACACCAAGAGAGCC 306  
QY 301 CTGGAACAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAACAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCCGACCAAGGCAAGGTGAGCGCAAGACTACCCCATCTGTGCAAACTGCAAGGCCAGATG 420  
DB 367 GCCGACCAAGGCAAGGTGAGCGCAAGACTACCCCATCTGTGCAAACTGCAAGGCCAGATG 426  
QY 421 GTGACCAAGGCGATCAGCCCGCGACCCCTGAACGCTGGGTGAAGGTGATGAGAGAGAG 480  
DB 427 GTGACCAAGGCGATCAGCCCGCGACCCCTGAACGCTGGGTGAAGGTGATGAGAGAGAG 486  
QY 481 GCCTTCAGCGCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 540  
DB 487 GCCTTCAGCGCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCGCCACCCCCAG 546  
QY 541 GACCTGAACAGATGTTGAACAACGTTGGCGGACCAAGGCCCGCATGAGATGCTGAAG 600  
DB 547 GACCTGAACAGATGTTGAACAACGTTGGCGGACCAAGGCCCGCATGAGATGCTGAAG 606  
QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCCC 660  
DB 607 GACACCATCAACGAGAGAGCGCGCGAGTGGGACCGCGTGCACCCCGTGCACGCGCGCCCC 666  
QY 661 ATCGCCCCCGGCGAGATGCGGAGCGCGCGCGGAGCGAGAGATCGCGCGGACCAAGCAACC 720  
DB 667 ATCGCCCCCGGCGAGATGCGGAGCGCGCGCGGAGCGAGAGATCGCGCGGACCAAGCAACC 726

QY	721	CTGCAGGAGCAGATCGCCTTGATGACCAAGCAACCCCCCATCCCGCTGGGCGCATCTAC	780
Db	727	CTGCAGGAGCAGATCGCCTGATGACCAAGCAACCCCCCATCCCGCTGGGCGCATCTAC	786
QY	781	AAGCGGTGATCATCTCTGGGCTTGAAACAAGATGTGCGGATGTACAGCCCCGTGAGCATC	840
Db	787	AAGCGGTGATCATCTCTGGGCTTGAAACAAGATGTGCGGATGTACAGCCCCGTGAGCATC	846
QY	841	CTGGAATCATCAAGCAGGGCCCCCAAGAGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAG	900
Db	847	CTGGAATCATCAAGCAGGGCCCCCAAGAGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAG	906
QY	901	ACCCTGCGCGCCGACAGAGCAACCCAGAGGTGAAGAACTGGATGACCGACACCTGCTG	960
Db	907	ACCCTGCGCGCCGACAGAGCAACCCAGAGGTGAAGAACTGGATGACCGACACCTGCTG	966
QY	961	GTCGAGAACGCCAACCCCGCATGCAAGACCATCTCGCGCTCTCGGCCCCGCGCCAGC	1020
Db	967	GTCGAGAACGCCAACCCCGCATGCAAGACCATCTCGCGCTCTCGGCCCCGCGCCAGC	1026
QY	1021	CTGGAAGAGATGATGACCGCCTGCCAGAGGCGCTGGGGCGGCCCAACCAAGGCCGCTG	1080
Db	1027	CTGGAAGAGATGATGACCGCCTGCCAGAGGCGCTGGGGCGGCCCAACCAAGGCCGCTG	1086
QY	1081	CTGGCCGAGGCGGATGAGCCAGGCCAACACCAAGCGTATGATGACGAAGAAGCAATTCAAG	1140
Db	1087	CTGGCCGAGGCGGATGAGCCAGGCCAACACCAAGCGTATGATGACGAAGAAGCAATTCAAG	1146
QY	1141	GGCCCCCGGCGCATCGTCAAGTGCCTTCAACTGGCGGCAAGAGGGCCACATCGCCGCAAC	1200
Db	1147	GGCCCCCGGCGCATCGTCAAGTGCCTTCAACTGGCGGCAAGAGGGCCACATCGCCGCAAC	1206
QY	1201	TGCGCGCGCCCCCGCAAGAGGCGCTGTGAAGTGCAGCAAGAGGGCCACAGATGAAG	1260
Db	1207	TGCGCGCGCCCCCGCAAGAGGCGCTGTGAAGTGCAGCAAGAGGGCCACAGATGAAG	1266
QY	1261	GACTGCACCCGAGCGCCAGGSCCACTTCTGGGCAAGATCTGGCCAGCCACAAGGGCCGC	1320
Db	1267	GACTGCACCCGAGCGCCAGGSCCACTTCTGGGCAAGATCTGGCCAGCCACAAGGGCCGC	1326
QY	1321	CCCCGCAACTTCTGACAGAGCGCCCCGAGGCCACCGCCCCCCCCCGCGAGAGCTTCCGC	1380
Db	1327	CCCCGCAACTTCTGACAGAGCGCCCCGAGGCCACCGCCCCCCCCCGCGAGAGCTTCCGC	1386
QY	1381	TTGAGAGAGACCAACCCCGGCCAAGAGCAGAGAGCAAGGACCGCGAGACCTGACCAGC	1440
Db	1387	TTGAGAGAGACCAACCCCGGCCAAGAGCAGAGAGCAAGGACCGCGAGACCTGACCAGC	1446
QY	1441	CTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAGTAA	1479
Db	1447	CTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAGTAA	1485

```

RESULT 5
US-10-190-435-19
; Sequence 19, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P/18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 4419

```

Query Match	Best Local Similarity	Score	DB	Length
99.9%	99.9%	1477.4	DB 15	4419
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
; TYPE: DNA				
; ORGANISM: Artificial Sequence				
; FEATURE:				
; OTHER INFORMATION: Description of Artificial Sequence: gagRtmuTatRevNeF_C				
US-10-190-435-19				
QY	1	ATGGGCGCCCGCGCCAGCATCTCTGCGCGCGCAAGCTGGACGCTGGGAGCGCATCCGC	60	
DB	7	ATGGGCGCCCGCGCCAGCATCTCTGCGCGCGCAAGCTGGACGCTGGGAGCGCATCCGC	66	
QY	61	CTGGCGCCCGCGCGCAAGAGTGTCTACATGATGAAGCACTGGTGTGGGCCAGCCGCGAG	120	
DB	67	CTGGCGCCCGCGCGCAAGAGTGTCTACATGATGAAGCACTGGTGTGGGCCAGCCGCGAG	126	
QY	121	CTGGAGAAAGTTTCGCTCTGAAACCCCGGCTGTGGAGACCAAGGAGGCTGGCAAGCATC	180	
DB	127	CTGGAGAAAGTTTCGCTCTGAAACCCCGGCTGTGGAGACCAAGGAGGCTGGCAAGCATC	186	
QY	181	ATCCGCCAGCTGCACCCCGCCCTGCAGACCGGCGAGCGAGAGCTGAAGAGCCTGTTCAAC	240	
DB	187	ATCCGCCAGCTGCACCCCGCCCTGCAGACCGGCGAGCGAGAGCTGAAGAGCCTGTTCAAC	246	
QY	241	ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGGTCGCGACACCAAGAGAGGCC	300	
DB	247	ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGGTCGCGACACCAAGAGAGGCC	306	
QY	301	CTGGACAAGATCGAGGAGGACGAACAAGTGCCAGCAGAGAATCCAGCAGGCCGAGGCC	360	
DB	307	CTGGACAAGATCGAGGAGGACGAACAAGTGCCAGCAGAGAATCCAGCAGGCCGAGGCC	366	
QY	361	GCCGACAAGGCGAAGGTGAGCCGAACCTACCCCATCTGTGAGAACTTGACAGGSCCAGATG	420	
DB	367	GCCGACAAGGCGAAGGTGAGCCGAACCTACCCCATCTGTGAGAACTTGACAGGSCCAGATG	426	
QY	421	GTCACACCAAGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAAAG	480	
DB	427	GTCACACCAAGCCATCAGCCCCCGCACCCCTGAACGCTGGGTGAAGGTGATCGAGAGAAAG	486	
QY	481	GCCTTCAGCCCGGAGGTGATCCCATGTTCACCGCCCTGACGAGGCGGCCACCCCCCAG	540	
DB	487	GCCTTCAGCCCGGAGGTGATCCCATGTTCACCGCCCTGACGAGGCGGCCACCCCCCAG	546	
QY	541	GACCTGAACAAGATGTTGAACACCGTGGGCGGCCACCAAGCCGCCATGAGATGCTGAAG	600	
DB	547	GACCTGAACAAGATGTTGAACACCGTGGGCGGCCACCAAGCCGCCATGAGATGCTGAAG	606	
QY	601	GACACCATCAACGAGGAGGCGCGCGAGTGGGAACGCGTGCACCCCGTGCACGCGCGCCCC	660	
DB	607	GACACCATCAACGAGGAGGCGCGCGAGTGGGAACGCGTGCACCCCGTGCACGCGCGCCCC	666	
QY	661	ATCGCCCCCGGCGAGATGCGCGGAGCCCGCGGCGAGCGACAATCGCCGCGACCAACAGCACC	720	
DB	667	ATCGCCCCCGGCGAGATGCGCGGAGCCCGCGGCGAGCGACAATCGCCGCGACCAACAGCACC	726	
QY	721	CTGCAGAGAGAGATCGCCTGAGATGACCAAGCAACCCGCCCATCCCGTGGGCGACATCTAC	780	
DB	727	CTGCAGAGAGAGATCGCCTGAGATGACCAAGCAACCCGCCCATCCCGTGGGCGACATCTAC	786	
QY	781	AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGTGAGCATC	840	
DB	787	AAGCGGTGATCATCTGGGCTGAACAAGATCGTGCGGATGTACAGCCCGTGAGCATC	846	
QY	841	CTGACATCAAGCAGGAGCCCAAGAGCCCTTCGCGCATACGTGACCGCTTCTTCAAG	900	
DB	847	CTGACATCAAGCAGGAGCCCAAGAGCCCTTCGCGCATACGTGACCGCTTCTTCAAG	906	
QY	901	ACCCTGCGCGCGAGCAGAGCAACCAAGAGGTGAAGAACTGATGACCGACACCCCTGCTG	960	
DB	907	ACCCTGCGCGCGAGCAGAGCAACCAAGAGGTGAAGAACTGATGACCGACACCCCTGCTG	966	

QY 961 GTGCAAGAACGCAACCCGAGCTGCAAGACCATCTTCCGCGCTCTTGGGCCCCGGGCGCCAGC 1020  
 DB 967 GTGCAAGAACGCAACCCGAGCTGCAAGACCATCTTCCGCGCTCTTGGGCCCCGGGCGCCAGC 1026  
 QY 1021 CTGAGAGGAGATGATGACCCGCTGCGAGGGCGGTGGGCGGGCCCCAGCCACAAAGGCCCGCTG 1080  
 DB 1027 CTGAGAGGAGATGATGACCCGCTGCGAGGGCGGTGGGCGGGCCCCAGCCACAAAGGCCCGCTG 1086  
 QY 1081 CTGAGCGAGGCGATGAGCCAGGCCCAACCAAGCGTGAATGATGCAAGAAAGCAACTTCAAG 1140  
 DB 1087 CTGAGCGAGGCGATGAGCCAGGCCCAACCAAGCGTGAATGATGCAAGAAAGCAACTTCAAG 1146  
 QY 1141 GGGCCCCGGGCGCATGCTCAAGTGTCTTCAACTGCGGCAAGAGGGCCCATCGCCCGCAAC 1200  
 DB 1147 GGGCCCCGGGCGCATGCTCAAGTGTCTTCAACTGCGGCAAGAGGGCCCATCGCCCGCAAC 1206  
 QY 1201 TGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCCAACAGATGAAG 1260  
 DB 1207 TGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCCAACAGATGAAG 1266  
 QY 1261 GACTGCAACGAGCGCGAGGCCCAACTTCTTGGGCAAGATCTGGCCCAAGAGGGCCCGC 1320  
 DB 1267 GACTGCAACGAGCGCGAGGCCCAACTTCTTGGGCAAGATCTGGCCCAAGAGGGCCCGC 1326  
 QY 1321 CCGCGCAACTTCTTGGGCAAGCGCGCGAGGCCCAAGAGGGCCCGCAGAGCTTCCGC 1380  
 DB 1327 CCGCGCAACTTCTTGGGCAAGCGCGCGAGGCCCAAGAGGGCCCGCAGAGCTTCCGC 1386  
 QY 1381 TTGAGAGAGACCAACCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 DB 1387 TTGAGAGAGACCAACCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
 QY 1441 CTGAAGAGCTGTTCGGCAACGACCCCTTGAAGCACTTAA 1479  
 DB 1447 CTGAAGAGCTGTTCGGCAACGACCCCTTGAAGCACTTAA 1485

RESULT 6  
 US-10-190-305A-14  
 / Sequence 14, Application US/10190305A  
 / Publication No. US20030198621A1  
 / GENERAL INFORMATION:  
 / APPLICANT: ZUR MESEDE, Jan  
 / APPLICANT: BARNETT, Susan  
 / APPLICANT: LIAN, Ying  
 / TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
 / TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
 / FILE REFERENCE: 2302-18702 / 18702.002  
 / CURRENT APPLICATION NUMBER: US/10/190,305A  
 / CURRENT FILING DATE: 2002-07-05  
 / NUMBER OF SEQ ID NOS: 93  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 14  
 / LENGTH: 4419  
 / TYPE: DNA  
 / ORGANISM: Artificial Sequence  
 / FEATURE:  
 / OTHER INFORMATION: Description of Artificial Sequence:  
 / OTHER INFORMATION: GagRTmutatRevNef\_C  
 / US-10-190-305A-14

Query Match 99.9%; Score 1477.4; DB 16; Length 4419;  
 Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
 Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGGGCGCGCGCGAGCATCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
 DB 7 ATGGGCGCGCGCGAGCATCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 66  
 QY 61 CTGGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGGCGAGCGCGAG 120  
 DB 67 CTGGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGGCGAGCGCGAG 126

QY 121 CTGAGAGATTGCGCTTGAACCCCGGCTGTGAGAGACGAGGGCTGCAAGCAGATC 180  
 DB 127 CTGAGAGATTGCGCTTGAACCCCGGCTGTGAGAGACGAGGGCTGCAAGCAGATC 186  
 QY 181 ATCCGCGAGTGCACCCCGCTTGCAGACCGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 DB 187 ATCCGCGAGTGCACCCCGCTTGCAGACCGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
 QY 241 ACCGTGGCCACCTGTACTGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
 DB 247 ACCGTGGCCACCTGTACTGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
 QY 301 CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 DB 307 CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
 QY 361 GCCGCAAGGGCAAGGTGAGCCAGAACTAACCCTATCGTGCAGAACTTGCAGAGGCGCAGATG 420  
 DB 367 GCCGCAAGGGCAAGGTGAGCCAGAACTAACCCTATCGTGCAGAACTTGCAGAGGCGCAGATG 426  
 QY 421 GTGCACCAAGGCATCAGCCCCCGCACCTTGAACGCTGGGTGAAGGTGATCGAGAGAGAG 480  
 DB 427 GTGCACCAAGGCATCAGCCCCCGCACCTTGAACGCTGGGTGAAGGTGATCGAGAGAGAG 486  
 QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGCCACCCCGCAG 540  
 DB 487 GCCTTCAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTGAGCGAGGGCGCCACCCCGCAG 546  
 QY 541 GACCTGAACAGATGTTGAACACCGTGGGCGGCCACCAAGCGCCCATGCAAGATGCTGAAG 600  
 DB 547 GACCTGAACAGATGTTGAACACCGTGGGCGGCCACCAAGCGCCCATGCAAGATGCTGAAG 606  
 QY 601 GACACCATCAACGAGAGAGAGCGCGCGAGTGGGACCGCGTGCAACCCCGTGCAAGCGCGGCCCC 660  
 DB 607 GACACCATCAACGAGAGAGCGCGCGAGTGGGACCGCGTGCAACCCCGTGCAAGCGCGGCCCC 666  
 QY 661 ATCGCCCCCGCGAGATGCGCGAGCCCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 DB 667 ATCGCCCCCGCGAGATGCGCGAGCCCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
 QY 721 CTGAGAGAGATGCGCTTGAACAGCAACCCCGCATCCCGTGGGCGCATCTAC 780  
 DB 727 CTGAGAGAGATGCGCTTGAACAGCAACCCCGCATCCCGTGGGCGCATCTAC 786  
 QY 781 AAGCGGTGATCATCTTGGGCTGAACAAAGATCGTGCGATGTAACGCCCGTGAGCATC 840  
 DB 787 AAGCGGTGATCATCTTGGGCTGAACAAAGATCGTGCGATGTAACGCCCGTGAGCATC 846  
 QY 841 CTGAGCATCAAGAGAGGGGCCCAAGAGAGCCCTCCGCACTACGTGAGACCGCTTCTTCAAG 900  
 DB 847 CTGAGCATCAAGAGAGGGGCCCAAGAGAGCCCTCCGCACTACGTGAGACCGCTTCTTCAAG 906  
 QY 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 DB 907 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966  
 QY 961 GTGAGAACGCGCAACCCCGCATGCAAGACCACTTGGCGGCTCTGCGGCCCGCGCGCAGC 1020  
 DB 967 GTGAGAACGCGCAACCCCGCATGCAAGACCACTTGGCGGCTCTGCGGCCCGCGCGCAGC 1026  
 QY 1021 CTGAGAGAGATGAGACCGCTTGCAGAGGGCGGCGCCAGCCACAAAGGCCCGCGTG 1080  
 DB 1027 CTGAGAGAGATGAGACCGCTTGCAGAGGGCGGCGCCAGCCACAAAGGCCCGCGTG 1086  
 QY 1081 CTGAGCGAGGCGATGAGCCAGGCCAACAAGCGTGAATGATGCAAGAAAGCAACTTCAAG 1140  
 DB 1087 CTGAGCGAGGCGATGAGCCAGGCCAACAAGCGTGAATGATGCAAGAAAGCAACTTCAAG 1146  
 QY 1141 GGGCCCCGGGCGCATGCTCAAGTGTCTTCAACTGCGGCAAGAGGGCCCATCGCCCGCAAC 1200  
 DB 1147 GGGCCCCGGGCGCATGCTCAAGTGTCTTCAACTGCGGCAAGAGGGCCCATCGCCCGCAAC 1206



QY 1441 CTGAAGAGCCTGTTCCGCAACGACCCCTGAGCCAGTAA 1479  
DB 4445 CTGAAGAGCCTGTTCCGCAACGACCCCTGAGCCATTA 4483

## RESULT 8

US-10-190-435-34  
; Sequence 34, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 34  
; LENGTH: 4606  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: gp160mod.TV1.dvl-gagmod.BW965  
US-10-190-435-34

Query Match 99.9%; Score 1477.4; DB 15; Length 4606;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;

Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCGCCGCGCGCATCTCTGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
DB 3128 ATGGGCGCCGCGCGCATCTCTGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 3187  
QY 61 CTGGCGCCCGCGCGCAAGTGTCTATGATGAAGCACTGTGTGGCGCAGCGCGAG 120  
DB 3188 CTGGCGCCCGCGCGCAAGTGTCTATGATGAAGCACTGTGTGGCGCAGCGCGAG 3247  
QY 121 CTGAAGAAGTTCGCTTGAACCCCGCTGCTGAGACCAAGCGGCTGCAAGCATC 180  
DB 3248 CTGAAGAAGTTCGCTTGAACCCCGCTGCTGAGACCAAGCGGCTGCAAGCATC 3307  
QY 181 ATCCGCGAGCTGACCCCGCTGAGACCGGCAAGGAGTGAAGAGCTGTTCAAC 240  
DB 3308 ATCCGCGAGCTGACCCCGCTGAGACCGGCAAGGAGTGAAGAGCTGTTCAAC 3367  
QY 241 ACCGTGGCCACCTTACTGCTGCAAGAGATCGAGTCCGCGACCAAGAGAGGCC 300  
DB 3368 ACCGTGGCCACCTTACTGCTGCAAGAGATCGAGTCCGCGACCAAGAGAGGCC 3427  
QY 301 CTGAACAAGATCGAGAGGAGAGCAAGTGCAGAGAGATCCAGCAGGCGCGAGGCC 360  
DB 3428 CTGAACAAGATCGAGAGGAGAGCAAGTGCAGAGAGATCCAGCAGGCGCGAGGCC 3487  
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACTAACCCTGTCAGAACTGCAAGGCGCAGATG 420  
DB 3488 GCCGACAAGGGCAAGGTGAGCCAGAACTAACCCTGTCAGAACTGCAAGGCGCAGATG 3547  
QY 421 GTGCACCAAGGCCATCAACCCCGCATCTGAACGCTTGGGTGAAGTGTATCGAGAGAA 480  
DB 3548 GTGCACCAAGGCCATCAACCCCGCATCTGAACGCTTGGGTGAAGTGTATCGAGAGAA 3607  
QY 481 GCCTTCAAGCCCGAGGTGATCCCATGTTTCAACCGCTTGAAGGAGGCGCCACCCCCAG 540  
DB 3608 GCCTTCAAGCCCGAGGTGATCCCATGTTTCAACCGCTTGAAGGAGGCGCCACCCCCAG 3667  
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCGCCCATGCAATGCTGAAG 600

DB 3668 GACCTGAACACGATGTTGAACACCGTGGCGGCCACCAAGCGCCCATGCAATGCTGAAG 3727  
QY 601 GACACCATCAACGAGAGGCGCGCGAGTGGACCGCGTGCAACCCCGTGACGCGCGGCC 660  
DB 3728 GACACCATCAACGAGAGGCGCGCGAGTGGACCGCGTGCAACCCCGTGACGCGCGGCC 3787  
QY 661 ATCGCCCCCGGCGAGATGCGCGAGCCCGCGCGAGCGACATCGCCGCGCACCAAGCACC 720  
DB 3788 ATCGCCCCCGGCGAGATGCGCGAGCCCGCGCGAGCGACATCGCCGCGCACCAAGCACC 3847  
QY 721 CTGCAGAGAGATCGCTGATGACCAAGACACCCCGCATCCCGTGCGGACATCTAC 780  
DB 3848 CTGCAGAGAGATCGCTGATGACCAAGACACCCCGCATCCCGTGCGGACATCTAC 3907  
QY 781 AAGCGGTGATCATCTGCGCTGAACCAAGATCGTGGATGTACAGCCCGTGAGCATC 840  
DB 3908 AAGCGGTGATCATCTGCGCTGAACCAAGATCGTGGATGTACAGCCCGTGAGCATC 3967  
QY 841 CTGACATCAAGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTTAAG 900  
DB 3968 CTGACATCAAGCAGGCGCCCAAGAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTTAAG 4027  
QY 901 ACCCTGCGCGCGAGAGAGACCCAGAGAGTGAAGACTGTGATGACCGACACCCCTGCTG 960  
DB 4028 ACCCTGCGCGCGAGAGAGACCCAGAGAGTGAAGACTGTGATGACCGACACCCCTGCTG 4087  
QY 961 GTGCAGAACCGCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCGCAGC 1020  
DB 4088 GTGCAGAACCGCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCCCGCGCGCAGC 4147  
QY 1021 CTGAGAGAGATGATGACCGCTGCGCAAGGCGCGCGCGCCAGCCACCAAGCGCGCTG 1080  
DB 4148 CTGAGAGAGATGATGACCGCTGCGCAAGGCGCGCGCGCCAGCCACCAAGCGCGCTG 4207  
QY 1081 CTGCGCCGAGCGCATGAGCCAGGCGCAACACCAAGCTGATGATGAGAAGCACTTCAAG 1140  
DB 4208 CTGCGCCGAGCGCATGAGCCAGGCGCAACACCAAGCTGATGATGAGAAGCACTTCAAG 4267  
QY 1141 GGGCCCCGGCGCATCTCTCAAGTCTTCACTGCGGCGAAGAGGCGCCACATCGCCGCAAC 1200  
DB 4268 GGGCCCCGGCGCATCTCTCAAGTCTTCACTGCGGCGAAGAGGCGCCACATCGCCGCAAC 4327  
QY 1201 TGGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGGGCAAGAGGCGCCACAGATGAAG 1260  
DB 4328 TGGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGGGCAAGAGGCGCCACAGATGAAG 4387  
QY 1261 GACTGACCGAGCGCCAGGCGCACTTCTGCGCAAGATCTGGGCCAGCCACCAAGGCGCGC 1320  
DB 4388 GACTGACCGAGCGCCAGGCGCACTTCTGCGCAAGATCTGGGCCAGCCACCAAGGCGCGC 4447  
QY 1321 CCGGCGCACTTCTGCAAGAGCGCGCGCGAGCGCCACCGCCCCCGCGGAGCTTCCGC 1380  
DB 4448 CCGGCGCACTTCTGCAAGAGCGCGCGCGAGCGCCACCGCCCCCGCGGAGCTTCCGC 4507  
QY 1381 TTGAGAGAGACCAACCCCGCGCAAGAGCAAGAGCAAGAGACCAAGAGCCCTGACCAAGC 1440  
DB 4508 TTGAGAGAGACCAACCCCGCGCAAGAGCAAGAGCAAGAGACCAAGAGCCCTGACCAAGC 4567  
QY 1441 CTGAAGAGCCTGTTCCGCAACGACCCCTGAGCCAGTAA 1479  
DB 4568 CTGAAGAGCCTGTTCCGCAACGACCCCTGAGCCATTA 4606

## RESULT 9

US-10-190-435-36  
; Sequence 36, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan

APPLICANT: VAN RENSBURG, Estrelita J.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
FILE REFERENCE: P18133.003 / 2302-18133  
CURRENT APPLICATION NUMBER: US/10/190,435  
CURRENT FILING DATE: 2002-12-30  
NUMBER OF SEQ ID NOS: 319  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 36  
LENGTH: 4615  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: gp160mod.TV1.dv2-gagmod.BW965  
US-10-190-435-36

Query Match 99.9%; Score 1477.4; DB 15; Length 4615;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCGAGCATCTCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
DB 3137 ATGGGCGCGCGCGCGAGCATCTCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 3196  
QY 61 CTGCGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGCACCTGGTGTGGCGCGCGCGAG 120  
DB 3197 CTGCGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGCACCTGGTGTGGCGCGCGAG 3256  
QY 121 CTGAGAGAGTTGCGCCCTGAAACCCCGGCTGTGAGAGCAAGCGGCTGCAAGCAGATC 180  
DB 3257 CTGAGAGAGTTGCGCCCTGAAACCCCGGCTGTGAGAGCAAGCGGCTGCAAGCAGATC 3316  
QY 181 ATCCGCGCAGCTGACACCCCGGCTGCAAGCGGCGCGAGAGCTGAAGAGCTGTCAAC 240  
DB 3317 ATCCGCGCAGCTGACACCCCGGCTGCAAGCGGCGCGAGAGCTGAAGAGCTGTCAAC 3376  
QY 241 ACCGTGGCGACCTGTACTGCGTGCAAGAGATCGAGTCCGCGACCAAGAGAGGCC 300  
DB 3377 ACCGTGGCGACCTGTACTGCGTGCAAGAGATCGAGTCCGCGACCAAGAGAGGCC 3436  
QY 301 CTGAGACAAGATCGAGAGAGCAGAACAGTGCAGAGATCCAGCAGCGCGAGGCC 360  
DB 3437 CTGAGACAAGATCGAGAGAGCAGAACAGTGCAGAGATCCAGCAGCGCGAGGCC 3496  
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACCTCCATCTGTGCAAACTTGCAAGGCGCAGATG 420  
DB 3497 GCCGACAAGGGCAAGGTGAGCCAGAACCTCCATCTGTGCAAACTTGCAAGGCGCAGATG 3556  
QY 421 GTGACACGAGCCATCAGCCCCCGCACCTGAAAGCGCTGGTGAAGGTGATCGAGAGAA 480  
DB 3557 GTGACACGAGCCATCAGCCCCCGCACCTGAAAGCGCTGGTGAAGGTGATCGAGAGAA 3616  
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCCAG 540  
DB 3617 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGGCGCCACCCCCAG 3676  
QY 541 GACCTGAACAAGATGTTGAACAAGTGGGCGCGCACAGCGCCCATGAGATGCTGAAG 600  
DB 3677 GACCTGAACAAGATGTTGAACAAGTGGGCGCGCACAGCGCCCATGAGATGCTGAAG 3736  
QY 601 GACACCATCAAGAGAGGCGCGCGAGTGGAGACCGGTGCAACCCGTGCAAGCGCGGCC 660  
DB 3737 GACACCATCAAGAGAGGCGCGCGAGTGGAGACCGGTGCAACCCGTGCAAGCGCGGCC 3796  
QY 661 ATGCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGCAGATCGCGCGCACAGCAGACC 720  
DB 3797 ATGCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGCAGATCGCGCGCACAGCAGACC 3856  
QY 721 CTGAGAGAGAGATCGCTGTGATGACCAAGAACCCCGCATCCCGTGGCGCAGATCTAC 780  
DB 3857 CTGAGAGAGAGATCGCTGTGATGACCAAGAACCCCGCATCCCGTGGCGCAGATCTAC 3916

QY 781 AAGCGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 840  
DB 3917 AAGCGTGATCATCTGGGCGCTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 3976  
QY 841 CTGACATCAAGCAGAGGCGCGCAAGAGCGCTTCCGCGCATACGTGAGCGCTTCTCAAG 900  
DB 3977 CTGACATCAAGCAGAGGCGCGCAAGAGCGCTTCCGCGCATACGTGAGCGCTTCTCAAG 4036  
QY 901 ACCCTGGCGCGCGAGCAGACACCCAGAGGTGAAGACTGATGACCGACACCTGCTG 960  
DB 4037 ACCCTGGCGCGCGAGCAGACACCCAGAGGTGAAGACTGATGACCGACACCTGCTG 4096  
QY 961 GTGAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCGCGCGCGCAGC 1020  
DB 4097 GTGAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCGCGCGCGCAGC 4156  
QY 1021 CTGAGAGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCGCGCGCGCGCGCGCTG 1080  
DB 4157 CTGAGAGAGATGATGACCGCTGCGCAGGCGGTGGCGCGCGCGCGCGCGCGCGCTG 4216  
QY 1081 CTGCGCGAGCGATGAGCGCAAGCCCAACACAGCGTGTATGATGAGAAAGCAACTTCAAG 1140  
DB 4217 CTGCGCGAGCGATGAGCGCAAGCCCAACACAGCGTGTATGATGAGAAAGCAACTTCAAG 4276  
QY 1141 GGCGCGCGCGCGCATCTGCAAGTGTCTCAACTGCGCAAGAGAGGCGCAACATCGCGCAAC 1200  
DB 4277 GGCGCGCGCGCGCATCTGCAAGTGTCTCAACTGCGCAAGAGAGGCGCAACATCGCGCAAC 4336  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCTGTGAAAGTGGCGCAAGAGAGGCGCAACAGATGAAG 1260  
DB 4337 TGCGCGCGCGCGCGCAAGAGGCTGTGAAAGTGGCGCAAGAGAGGCGCAACAGATGAAG 4396  
QY 1261 GACTGCAACGAGCGCGCAAGCTTCTGGCAAGATCTGGCCAGCCCAAGAGGCGCGC 1320  
DB 4397 GACTGCAACGAGCGCGCAAGCTTCTGGCAAGATCTGGCCAGCCCAAGAGGCGCGC 4456  
QY 1321 CCCGCAACTTCTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1380  
DB 4457 CCCGCAACTTCTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4516  
QY 1381 TTGAGAGAGACACCCCGCGCGCAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 4517 TTGAGAGAGACACCCCGCGCGCGCAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 4576  
QY 1441 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAGTAA 1479  
DB 4577 CTGAAGAGCTGTTCGCAACGACCCCTGAGCCAGTAA 4615

RESULT 10  
US-10-190-435-38  
Sequence 38, Application US/10190435  
Publication No. US20030143248A1  
GENERAL INFORMATION:  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: BARNETT, Susan W.  
APPLICANT: LIAN, Ying  
APPLICANT: ENGELBRECHT, Susan  
APPLICANT: VAN RENSBURG, Estrelita J.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
FILE REFERENCE: P18133.003 / 2302-18133  
CURRENT APPLICATION NUMBER: US/10/190,435  
CURRENT FILING DATE: 2002-12-30  
NUMBER OF SEQ ID NOS: 319  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 38  
LENGTH: 4702  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: gp160mod.TV1-gagmod.BW965

US-10-190-435-38

Query Match 99.9%; Score 1477.4; DB 15; Length 4702;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGGCGCCGCGCCAGCATCTGCGGGCGGCAAGCTGGACGCGCTGGGAGCGCATCCGC 60
Db 3224 ATGGGCGCCGCGCCAGCATCTGCGGGCGGCAAGCTGGACGCGCTGGGAGCGCATCCGC 3283
QY 61 CTGCGCGCGCGCGGCAAGAGTGCTACATGATGAAGCACTGTGTGGGCCAGCGCGGAG 120
Db 3284 CTGCGCGCGCGCGGCAAGAGTGCTACATGATGAAGCACTGTGTGGGCCAGCGCGGAG 3343
QY 121 CTGGAAGAAGTTGCGCCCTGAAACCCCGGCTGTGAGACCAAGCGGCTGCAAGCAGATC 180
Db 3344 CTGGAAGAAGTTGCGCCCTGAAACCCCGGCTGTGAGACCAAGCGGCTGCAAGCAGATC 3403
QY 181 ATCCGCGCAGCTGCACCCCGCGCTGCAAGCCGGCAGGAGAGCTGAAGAGCTGTTCAC 240
Db 3404 ATCCGCGCAGCTGCACCCCGCGCTGCAAGCCGGCAGGAGAGCTGAAGAGCTGTTCAC 3463
QY 241 ACCGTGGCCACCCCTGTAATGCGTGCAAGAAAGTCCGCGCAACCAAGGAGGCC 300
Db 3464 ACCGTGGCCACCCCTGTAATGCGTGCAAGAAAGTCCGCGCAACCAAGGAGGCC 3523
QY 301 CTGGAACAAGATCGAGAGGAGCAGAAAGTGCACAGAAATCCAGAGGCCGAGGCC 360
Db 3524 CTGGAACAAGATCGAGAGGAGCAGAAAGTGCACAGAAATCCAGAGGCCGAGGCC 3583
QY 361 GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATGTCAGAACTGCAAGGGCCAGATG 420
Db 3584 GCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATGTCAGAACTGCAAGGGCCAGATG 3643
QY 421 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAACGCTGGGTGAAGGTGATCGAGAGAAG 480
Db 3644 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAACGCTGGGTGAAGGTGATCGAGAGAAG 3703
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCCACCCCGCAG 540
Db 3704 GCCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGGCCACCCCGCAG 3763
QY 541 GACCTGAACACGATGTTGAACACCGTGGGGCGGCCACAGGCCGCCATGACAGATGCTGAAG 600
Db 3764 GACCTGAACACGATGTTGAACACCGTGGGGCGGCCACAGGCCGCCATGACAGATGCTGAAG 3823
QY 601 GACACCATCAACGAGAGGGCGCGCGAGTGGAGCCGCGTGCACCCCGTGCACGCGCGGCC 660
Db 3824 GACACCATCAACGAGAGGGCGCGCGAGTGGAGCCGCGTGCACCCCGTGCACGCGCGGCC 3883
QY 661 ATCGCCCCCGGCGCAGATGCGCGAGCCCGCGGCGAGGACATGCGCGGCAACCAAGCACC 720
Db 3884 ATCGCCCCCGGCGCAGATGCGCGAGCCCGCGGCGAGGACATGCGCGGCAACCAAGCACC 3943
QY 721 CTGCAAGGACAGATGCGCTGATGACCAACACCCCCCATCCCGTGGGGCAGCATCTAC 780
Db 3944 CTGCAAGGACAGATGCGCTGATGACCAACACCCCCCATCCCGTGGGGCAGCATCTAC 4003
QY 781 AAGCGGTGATCATCTGGGCTGMAACAAGATGTCGGGATGTACAGCCCCGTGAGCATC 840
Db 4004 AAGCGGTGATCATCTGGGCTGMAACAAGATGTCGGGATGTACAGCCCCGTGAGCATC 4063
QY 841 CTGACATCAAGCAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 900
Db 4064 CTGACATCAAGCAGGGCCCCCAAGAGCCCTTCCGCGACTACGTGACCGCTTCTTCAAG 4123
QY 901 ACCCTGCGCGCGGAGAGACACCAAGAGTGAAGAACTGGATGACGACACCTGCTG 960
Db 4124 ACCCTGCGCGCGGAGAGACACCAAGAGTGAAGAACTGGATGACGACACCTGCTG 4183
QY 961 GTGCAAGACGCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCCGCGCGCAGC 1020
Db 4184 GTGCAAGACGCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGCCCCGCGCGCAGC 4243
```

```
QY 1021 CTGAGAGAGATGATGACCGCTGCGCAAGGGCGGTGGGCGGCCCAAGCCACAAGGCCGCGTG 1080
Db 4244 CTGAGAGAGATGATGACCGCTGCGCAAGGGCGGTGGGCGGCCCAAGCCACAAGGCCGCGTG 4303
QY 1081 CTGGCCGAGGCGATGAGCCAGGCGCAACACCAAGCGTGTATGATGCAAGAGCACTTCAAG 1140
Db 4304 CTGGCCGAGGCGATGAGCCAGGCGCAACACCAAGCGTGTATGATGCAAGAGCACTTCAAG 4363
QY 1141 GGGCCCCGGCGCATGCTCAAGTGTCTCAACTGCGGCAAGGAGGCCACATGCGCGCAAC 1200
Db 4364 GGGCCCCGGCGCATGCTCAAGTGTCTCAACTGCGGCAAGGAGGCCACATGCGCGCAAC 4423
QY 1201 TGCGCGCGCCCCCGCAAGAGGCTGCTGAAAGTCCGGCAAGGAGGCCACCAAGATGAAG 1260
Db 4424 TGCGCGCGCCCCCGCAAGAGGCTGCTGAAAGTCCGGCAAGGAGGCCACCAAGATGAAG 4483
QY 1261 GACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCAGCCACAAGGGCGCG 1320
Db 4484 GACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCAGCCACAAGGGCGCG 4543
QY 1321 CCGGCAACTTCTGCAAGAGCGCGCCCGAGCCACCGCCCGCGCGGAGAGCTTCCGC 1380
Db 4544 CCGGCAACTTCTGCAAGAGCGCGCCCGAGCCACCGCCCGCGCGGAGAGCTTCCGC 4603
QY 1381 TTGAGAGAGACCAACCCCGCGCAAGAGCAGAGAGCAAGAGCAAGGAGCCCTGACCAAGC 1440
Db 4604 TTGAGAGAGACCAACCCCGCGCAAGAGCAGAGAGCAAGAGCAAGGAGCCCTGACCAAGC 4663
QY 1441 CTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAGTAA 1479
Db 4664 CTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAGTAA 4702
```

## RESULT 11

US-10-190-435-17

; Sequence 17, Application US/10190435

; Publication No. US20030143248A1

; GENERAL INFORMATION:

; APPLICANT: ZUR MEGEDE, Jan

; APPLICANT: BARNETT, Susan W.

; APPLICANT: LIAN, Ying

; APPLICANT: VAN RENSBURG, Susan

; TITLE OF INVENTION: ENGELBRECHT, Estrelita J.

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C

; FILE REFERENCE: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

; CURRENT FILING DATE: 2002-12-30

; CURRENT FILING DATE: 2002-12-30

; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 17

; LENGTH: 4716

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:

; OTHER INFORMATION: GagProteinMultatRevNef\_C

US-10-190-435-17

Query Match 99.9%; Score 1477.4; DB 15; Length 4716;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGCGCCGCGCCAGCATCTGCGCGGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60
Db 7 ATGGGCGCCGCGCCAGCATCTGCGCGGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 66
QY 61 CTGCGCGCGCGCGCAAGAGTGTACTATGATGAAGACACCTGTGTGGCCAGCGCGAG 120
Db 67 CTGCGCGCGCGCGCAAGAGTGTACTATGATGAAGACACCTGTGTGGCCAGCGCGAG 126
QY 121 CTGGAAGAAGTTGCGCCCTGAAACCCCGGCTGCTGGAAGACGAGAGGGCTGCAAGCAGATC 180
```

Db 127 CTGAGAA GTTCG CCGCTGA ACCCGGCTGTGAGACCGAGCGGCTGCAAGCAGATC 186  
QY 181 ATCCGCA GCTGCA CCGCCCTG CAGACCGG CAGCGAGCTGAAGCCTGTTCAAC 240  
Db 187 ATCCGCA GCTGCA CCGCCCTG CAGACCGG CAGCGAGCTGAAGCCTGTTCAAC 246  
QY 241 ACCGTGG CCACTCTG TACTGCTG CACGAGAA GATCGAGTCCGCGACCAAGAGGCC 300  
Db 247 ACCGTGG CCACTCTG TACTGCTG CACGAGAA GATCGAGTCCGCGACCAAGAGGCC 306  
QY 301 CTGAGCA AGATCGAG GAGAGCAG AAAGTGCC AGACAGAA GATCCAGAGGCCGAGGCC 360  
Db 307 CTGAGCA AGATCGAG GAGAGCAG AAAGTGCC AGACAGAA GATCCAGAGGCCGAGGCC 366  
QY 361 GCCGACA GGGGCA AGGTGAG CCGAAGTA CCCCATCG TGACAGAA CCGAGGGCCAGATG 420  
Db 367 GCCGACA GGGGCA AGGTGAG CCGAAGTA CCCCATCG TGACAGAA CCGAGGGCCAGATG 426  
QY 421 GTGACAC GAGCCATC AGCCCCG CACCCCTG AAGCGCTG GGTGAAGTGATCGAGAGAG 480  
Db 427 GTGACAC GAGCCATC AGCCCCG CACCCCTG AAGCGCTG AAGTGATCGAGAGAGAG 486  
QY 481 GCCTTCA GCGCCCGA GGTGAT CCCCATTG TCAACG CCGCTGA GCGAGGGGCCACCCCCAG 540  
Db 487 GCCTTCA GCGCCCGA GGTGAT CCCCATTG TCAACG CCGCTGA GCGAGGGGCCACCCCCAG 546  
QY 541 GACCTGA AACACGAT GTTGAAC ACCGTGG CGGGCC ACGAGGCC CCGCATG CAGATGCTGAAG 600  
Db 547 GACCTGA AACACGAT GTTGAAC ACCGTGG CGGGCC ACGAGGCC CCGCATG CAGATGCTGAAG 606  
QY 601 GACACCA TCACGAG GAGAGG CCGCGG CAGGTGG GAGCCG CGGTGCA CCGCGTGCACGCGGCCCC 660  
Db 607 GACACCA TCACGAG GAGAGG CCGCGG CAGGTGG GAGCCG CGGTGCA CCGCGTGCACGCGGCCCC 666  
QY 661 ATCGCCC CCGGCG CAGATG CGCGAG CCGCGG CCGGCA CGACATG CCGGCA CCAACGACACC 720  
Db 667 ATCGCCC CCGGCG CAGATG CGCGAG CCGCGG CCGGCA CGACATG CCGGCA CCAACGACACC 726  
QY 721 CTGCAGA GACAGAT CGCCTG GATGAC CAGACAA CCCCCC ATCCCG TGGGCG CAGATCTAAC 780  
Db 727 CTGCAGA GACAGAT CGCCTG GATGAC CAGACAA CCCCCC ATCCCG TGGGCG CAGATCTAAC 786  
QY 781 AAGCGGT GATCAT CCTTGG GGCCTG AACAAG ATCGTG CGGATGTACAGCCCCGTGAGCATC 840  
Db 787 AAGCGGT GATCAT CCTTGG GGCCTG AACAAG ATCGTG CGGATGTACAGCCCCGTGAGCATC 846  
QY 841 CTGGAAT CATGAAG CAGGGG CCCCCA AGGAGC CTTCCG CGACTA CGTGGA CCGCTTCTTCAAG 900  
Db 847 CTGGAAT CATGAAG CAGGGG CCCCCA AGGAGC CTTCCG CGACTA CGTGGA CCGCTTCTTCAAG 906  
QY 901 ACCCTG CGCGCG CGAGCAG ACGACCC CAGAGGTGA AGAACTG ATGACCGA CACCCCTGCTG 960  
Db 907 ACCCTG CGCGCG CGAGCAG ACGACCC CAGAGGTGA AGAACTG ATGACCGA CACCCCTGCTG 966  
QY 961 GTGCAGA ACGCCA ACCCGG ACTGCA AGACCAT CTTGCG CGCTCTCG GCGCCCGGCGCCAGC 1020  
Db 967 GTGCAGA ACGCCA ACCCGG ACTGCA AGACCAT CTTGCG CGCTCTCG GCGCCCGGCGCCAGC 1026  
QY 1021 CTGAGGA GAGATGAT GACCGC CTGCA GGGGCG GTGGCG GCGCCCA GCAAGGCGCGGCTG 1080  
Db 1027 CTGAGGA GAGATGAT GACCGC CTGCA GGGGCG GTGGCG GCGCCCA GCAAGGCGCGGCTG 1086  
QY 1081 CTGGCG GAGGGCG ATGAGC CAGGCC AACAAC CAGCGTGA TGATGCA GAAAGCAACTTCAAG 1140  
Db 1087 CTGGCG GAGGGCG ATGAGC CAGGCC AACAAC CAGCGTGA TGATGCA GAAAGCAACTTCAAG 1146  
QY 1141 GGGCCCC GGGCGCAT CGTCA AGTGCTTCA ACTGCG GCAAGGAG GGGCCA CATGCGCCGCAAC 1200  
Db 1147 GGGCCCC GGGCGCAT CGTCA AGTGCTTCA ACTGCG GCAAGGAG GGGCCA CATGCGCCGCAAC 1206  
QY 1201 TGCCGCG CCCCCG CAAAGG GGTGCTG GAAAGTG CGGCAAGGAG GGGCCA CAGATGAAG 1260

Db 1207 TGCCGCG CCCCCG CAAAGG GGTGCTG GAAAGTG CGGCAAGGAG GGGCCA CAGATGAAG 1266  
QY 1261 GACTGAC CCGAGCG CGGCAAG CTTCTG GGCAGAA GTCTGG CCGCAG CCAAGGGCCGC 1320  
Db 1267 GACTGAC CCGAGCG CGGCAAG CTTCTG GGCAGAA GTCTGG CCGCAG CCAAGGGCCGC 1326  
QY 1321 CCGGCA ACTTCTG CAGAG CCGCGG CAGAGCC CCGCAG CCGCCCC CCGCGAGAGCTTCCGC 1380  
Db 1327 CCGGCA ACTTCTG CAGAG CCGCGG CAGAGCC CCGCAG CCGCCCC CCGCGAGAGCTTCCGC 1386  
QY 1381 TTGAGAG AACCAC CCGCGG CAGAGCA GAGAGCA GAGAGG ACCCGCAG ACCCTGACCAAC 1440  
Db 1387 TTGAGAG AACCAC CCGCGG CAGAGCA GAGAGCA GAGAGG ACCCGCAG ACCCTGACCAAC 1446  
QY 1441 CTGAAG AGCCTGT TCGCAAC GACCCCTG AGCCAGTAA 1479  
Db 1447 CTGAAG AGCCTGT TCGCAAC GACCCCTG AGCCAGAAA 1485

## RESULT 12

US-10-190-305A-13  
; Sequence 13, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190,305A  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 4716  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; US-10-190-305A-13

Query Match 99.9%; Score 1477.4; DB 16; Length 4716;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCG CCGCGCG CAGCAT CCTGCG CGGGCG GCAAGCTG GACCGC CTGGGAGCGCATCCGC 60  
Db 7 ATGGGCG CCGCGCG CAGCAT CCTGCG CGGGCG GCAAGCTG GACCGC CTGGGAGCGCATCCGC 66  
QY 61 CTGGGCG CCGGGCG GCAAGGTGCTA CATGATGA AGCACTGTG TGGGGCCAGCCGCGAG 120  
Db 67 CTGGGCG CCGGGCG GCAAGGTGCTA CATGATGA AGCACTGTG TGGGGCCAGCCGCGAG 126  
QY 121 CTGAGAA GTTCG CCGCTGA ACCCGG CCGCTG TGAGAA CCAAGGGGCTGCAAGCAGATC 180  
Db 127 CTGAGAA GTTCG CCGCTGA ACCCGG CCGCTG TGAGAA CCAAGGGGCTGCAAGCAGATC 186  
QY 181 ATCCGCC AGCTGCA CCGCGC CTGCA GACCGG CAGCGAGAGCTGA AGAGCCTGTTCAAC 240  
Db 187 ATCCGCC AGCTGCA CCGCGC CTGCA GACCGG CAGCGAGAGCTGA AGAGCCTGTTCAAC 246  
QY 241 ACCGTGG CCACTGTA CTGCGTGA CGAAGATG AGGTCCG GACA CCAAGAGGGCC 300  
Db 247 ACCGTGG CCACTGTA CTGCGTGA CGAAGATG AGGTCCG GACA CCAAGAGGGCC 306  
QY 301 CTGACA AAGATCGAG GAGAGCAG AAAGTGCC AGCAGAA GATCCA GAGGCCGAGGCC 360  
Db 307 CTGACA AAGATCGAG GAGAGCAG AAAGTGCC AGCAGAA GATCCA GAGGCCGAGGCC 366  
QY 361 GCCGACA GGGGCA AGGTGAG CCGAAGTA CCCCCTAT CGTGCA GAACTG CAGGGGCCAGATG 420

```
Db 367 GCCGACAAAGGGCAAGGTGAGCCAGAACTACCCCATCGTGAGAACTTGACAGGCGCCAGATG 426
QY 421 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAACGCTGGGTGAAGGTGATCGAGAGAAG 480
Db 427 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAACGCTGGGTGAAGGTGATCGAGAGAAG 486
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCCAG 540
Db 487 GCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCCAG 546
QY 541 GACCTGAAACAGATGTTGAACACCGTGGGCGGCCACCAAGGCGCCCATGAGATGCTGAAG 600
Db 547 GACCTGAAACAGATGTTGAACACCGTGGGCGGCCACCAAGGCGCCCATGAGATGCTGAAG 606
QY 601 GACACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCTGCACCCCGTGCACGCGCGCCC 660
Db 607 GACACCATCAACGAGAGAGGCGCGCGAGTGGGACCGCTGCACCCCGTGCACGCGCGCCC 666
QY 661 ATGCCCCCGGCGCAGATGCGCGGAGCCCCGCGGAGCGACATCGCGCGGACCAACGACACC 720
Db 667 ATGCCCCCGGCGCAGATGCGCGGAGCCCCGCGGAGCGACATCGCGCGGACCAACGACACC 726
QY 721 CTGCAGGAGCAGATCGCTGTGATGACCAACAACCCCATCCCGTGGGCGACATCTAC 780
Db 727 CTGCAGGAGCAGATCGCTGTGATGACCAACAACCCCATCCCGTGGGCGACATCTAC 786
QY 781 AAGCGGTGATCATCTGGGCTGTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 840
Db 787 AAGCGGTGATCATCTGGGCTGTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC 846
QY 841 CTGCACATCAAGCAGGCGCCCAAGGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900
Db 847 CTGCACATCAAGCAGGCGCCCAAGGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 906
QY 901 ACCCTGCGCGCGGAGCAGACCAACGAGGTGAAGAACTGATGACGACACCTGTCTG 960
Db 907 ACCCTGCGCGCGGAGCAGACCAACGAGGTGAAGAACTGATGACGACACCTGTCTG 966
QY 961 GTGCAGAAAGCCAAACCCGACTGCAGAACCATCTGCGCGCTCTGCGCCCCGCGCGAGC 1020
Db 967 GTGCAGAAAGCCAAACCCGACTGCAGAACCATCTGCGCGCTCTGCGCCCCGCGCGAGC 1026
QY 1021 CTGGAAGAGATGATGACCGCTGCGAGGGCGTGGGCGGCCCAAGGCCCGCGCTG 1080
Db 1027 CTGGAAGAGATGATGACCGCTGCGAGGGCGTGGGCGGCCCAAGGCCCGCGCTG 1086
QY 1081 CTGGCCGAGGCGATGAGCCAGGCCAACACACAGCGTATGATGACAGAGCAACTTCAAG 1140
Db 1087 CTGGCCGAGGCGATGAGCCAGGCCAACACACAGCGTATGATGACAGAGCAACTTCAAG 1146
QY 1141 GGGCCCCGGCGCATCTCAAGTGTCTTCAACTGCGGCAAGAGGGGCCCATGCGCCGCAAC 1200
Db 1147 GGGCCCCGGCGCATCTCAAGTGTCTTCAACTGCGGCAAGAGGGGCCCATGCGCCGCAAC 1206
QY 1201 TGCGCGCGCCCCCGCAAGAAAGGCTGTGGAAGTGGCGCAAGGAGGCCACCAAGTGAAG 1260
Db 1207 TGCGCGCGCCCCCGCAAGAAAGGCTGTGGAAGTGGCGCAAGGAGGCCACCAAGTGAAG 1266
QY 1261 GACTGCACCGAGCGCGCAAGGCCAACTTCTTGCGCAAGATCTGCGCCACCAAGGCGCGC 1320
Db 1267 GACTGCACCGAGCGCGCAAGGCCAACTTCTTGCGCAAGATCTGCGCCACCAAGGCGCGC 1326
QY 1321 CCGGCAACTTCTTGCAAGAGCGCGCGAGGCCACCGCCCCCGCGCGAGAGCTTCCGC 1380
Db 1327 CCGGCAACTTCTTGCAAGAGCGCGCGAGGCCACCGCCCCCGCGCGAGAGCTTCCGC 1386
QY 1381 TTGAGAGAGACCAACCCCGGCAAGAGCAGAGAGCAAGAACCGGAGACCTTGACAGC 1440
Db 1387 TTGAGAGAGACCAACCCCGGCAAGAGCAGAGAGCAAGAACCGGAGACCTTGACAGC 1446
QY 1441 CTGAAGAGCCTGTTGGCAACGACCCCTGAGCCAGTAA 1479
Db 1447 CTGAAGAGCCTGTTGGCAACGACCCCTGAGCCAGAAA 1485
```

```
RESULT 13
US-10-190-435-20
; Sequence 20, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2742
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagTatRevNef_C
US-10-190-435-20
```

```
Query Match 99.8%; Score 1475.8; DB 15; Length 2742;
Best Local Similarity 99.9%; Pred. No. 2e-307;
Matches 1477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 ATGGGGCGCCCGCCAGCATCTCTGCGCGCGGCGCAAGCTGAGCCCTGGAGCGCATCCGC 60
Db 7 ATGGGGCGCCCGCCAGCATCTCTGCGCGCGGCGCAAGCTGAGCCCTGGAGCGCATCCGC 66
QY 61 CTGGCGCCCGGCGCAAGAGTGTACATGATGAAGACACCTGTGTGGCCAGCCGCGAG 120
Db 67 CTGGCGCCCGGCGCAAGAGTGTACATGATGAAGACACCTGTGTGGCCAGCCGCGAG 126
QY 121 CTGGAAGAATTGCCCCCTGAACCCCGGCTGTGGAAGACCAAGGAGGCTGCAAGCAGATC 180
Db 127 CTGGAAGAATTGCCCCCTGAACCCCGGCTGTGGAAGACCAAGGAGGCTGCAAGCAGATC 186
QY 181 ATCCGCCAGCTGACACCCCGCTGCAGACCGGCAAGGAGCTGAAGAGCTGTTCAC 240
Db 187 ATCCGCCAGCTGACACCCCGCTGCAGACCGGCAAGGAGCTGAAGAGCTGTTCAC 246
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGAGAGATCGAGTCCGCGACACCAAGAGGCC 300
Db 247 ACCGTGGCCACCTGTACTGCGTGCAGAGAGATCGAGTCCGCGACACCAAGAGGCC 306
QY 301 CTGGAACAAGATGAGAGAGAGCAACAAGTCCAGAGAGATCCAGAGGCCGAGGCC 360
Db 307 CTGGAACAAGATGAGAGAGAGCAACAAGTCCAGAGAGATCCAGAGGCCGAGGCC 366
QY 361 GCCGACAAAGGCAAGGTGAGGCCGAACCTACCCCATCTGTCAGAACTTGAGGGCCAGATG 420
Db 367 GCCGACAAAGGCAAGGTGAGGCCGAACCTACCCCATCTGTCAGAACTTGAGGGCCAGATG 426
QY 421 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAACGCTGGGTGAAGGTGATCGAGAGAAG 480
Db 427 GTGCACCAAGGCCATCAGCCCCCGCACCTGAAACGCTGGGTGAAGGTGATCGAGAGAAG 486
QY 481 GCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCCAG 540
Db 487 GCCTTCAGCCCCGAGGTGATCCCATGTTCACCGCCCTGAGCGAGGCGCCACCCCCAG 546
QY 541 GACCTGAACAGATGTTGAACACCGTGGGCGGCCACCAAGGCGCCCATGAGATGCTGAAG 600
Db 547 GACCTGAACAGATGTTGAACACCGTGGGCGGCCACCAAGGCGCCCATGAGATGCTGAAG 606
QY 601 GACACCATCAACGAGAGGCGCGGAGTGGACCGCGTGACCCCGTGACCGCGCGCCC 660
```

Db	607	GACACCATCAACGAGGAGGCGCGGAGTGGGACCGCGTGCAACCCCGTGCAACGCCGGCCCC	666
Qy	661	ATGCCCCCGGCCAGATGCGCGAGCCCCCGCGCGAGCGACATCGCCGGCACCAAGCACCC	720
Db	667	ATGCCCCCGGCCAGATGCGCGAGCCCCCGCGCGAGCGACATCGCCGGCACCAAGCACCC	726
Qy	721	CTGCAAGGAGCAGATCGCCTGGATGACCAAGCAACCCCCCATCCCGTGGGCGCATCTAC	780
Db	727	CTGCAAGGAGCAGATCGCCTGGATGACCAAGCAACCCCCCATCCCGTGGGCGCATCTAC	786
Qy	781	AAAGCGGTGATCATCTCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC	840
Db	787	AAAGCGGTGATCATCTCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCCGTGAGCATC	846
Qy	841	CTGGAATCAAGCAGGAGGCCCCCAAGAGCCCTTCCGCGACTACGTGGAACCGCTTCTCAAG	900
Db	847	CTGGAATCAAGCAGGAGGCCCCCAAGAGCCCTTCCGCGACTACGTGGAACCGCTTCTCAAG	906
Qy	901	ACCCTGCGCGCCGAGCAGAGCACCCAGAGGTTGAAGAACTGATGACCGACACCCCTGCTG	960
Db	907	ACCCTGCGCGCCGAGCAGAGCACCCAGAGGTTGAAGAACTGATGACCGACACCCCTGCTG	966
Qy	961	GTGCAGAACGCCAACCCCGACTGTCAGAACCATCTGCGGCTCTCGGCCCCGGCGGCAGC	1020
Db	967	GTGCAGAACGCCAACCCCGACTGTCAGAACCATCTGCGGCTCTCGGCCCCGGCGGCAGC	1026
Qy	1021	CTGGAGGAGATGATGACCGGCTGTCAGGGCGTGGGGCGGCCCCAGCCACAAAGGCCGCGTG	1080
Db	1027	CTGGAGGAGATGATGACCGGCTGTCAGGGCGTGGGGCGGCCCCAGCCACAAAGGCCGCGTG	1086
Qy	1081	CTGGCCGAGGCGGATGAGCCAGGCCAACACCAAGCTGATGATGACGAAGAGCAACTTCAAG	1140
Db	1087	CTGGCCGAGGCGGATGAGCCAGGCCAACACCAAGCTGATGATGACGAAGAGCAACTTCAAG	1146
Qy	1141	GGCCCCCGGCGCATCGTCAAGTGTCTCAACTGCGGCCAAGGAGGCCACATCGCCGCAAC	1200
Db	1147	GGCCCCCGGCGCATCGTCAAGTGTCTCAACTGCGGCCAAGGAGGCCACATCGCCGCAAC	1206
Qy	1201	TGCGCGCGCCCCCGCAAGAAAGGCTGCTGGAAGTGCGGCAAGGAGGGCCACAGATGAAG	1260
Db	1207	TGCGCGCGCCCCCGCAAGAAAGGCTGCTGGAAGTGCGGCAAGGAGGGCCACAGATGAAG	1266
Qy	1261	GACTGCACCCGAGCGCCAGGCCAACTTCTTGGGCAAGATCTGGGCCAGCCACAAGGGCCGC	1320
Db	1267	GACTGCACCCGAGCGCCAGGCCAACTTCTTGGGCAAGATCTGGGCCAGCCACAAGGGCCGC	1326
Qy	1321	CCCGGCAACTTCTCTGACAGGCGCGCCCGAGCCACCGCCCCCGCGCGAGAGCTTCCGC	1380
Db	1327	CCCGGCAACTTCTCTGACAGGCGCGCCCGAGCCACCGCCCCCGCGCGAGAGCTTCCGC	1386
Qy	1381	TTGAGAGAGACCAACCCCGGCCAGAAAGCAGAGAGCAAGGACCGCGAGACCTGACCCAGC	1440
Db	1387	TTGAGAGAGACCAACCCCGGCCAGAAAGCAGAGAGCAAGGACCGCGAGACCTGACCCAGC	1446
Qy	1441	CTGAAGAGCCTGTTGGCAACGACCCCTGAGCCAGTAA 1479	
Db	1447	CTGAAGAGCCTGTTGGCAACGACCCCTGAGCCAGTAA 1485	

```

RESULT 14
US-10-190-305A-15
; Sequence 15, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93

```

Query	Match	Similarity	Score	DB	Length
1	ATGGGCGCCCGCCAGCATCTCGCGCGCGGCAAGCTGGACGCTGGAGCGCATCCGC	99.8%	1475.8	16	2742
7	ATGGGCGCCCGCCAGCATCTCGCGCGCGGCAAGCTGGACGCTGGAGCGCATCCGC	99.8%	1475.8	16	2742
61	CTGGCGCCCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGCCAGCCGCGAG	99.8%	1475.8	16	2742
67	CTGGCGCCCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGCCAGCCGCGAG	99.8%	1475.8	16	2742
121	CTGGAGAGAGTTCCGCTGAAACCCCGGCTGTGGAGACCGAGGCGCTGCAAGCATC	99.8%	1475.8	16	2742
127	CTGGAGAGAGTTCCGCTGAAACCCCGGCTGTGGAGACCGAGGCGCTGCAAGCATC	99.8%	1475.8	16	2742
181	ATCCGCGCAGCTGCAACCCCGCTGCAAGCCGCGAGCGAGAGCTGAAGAGCTGTTCAAC	99.8%	1475.8	16	2742
187	ATCCGCGCAGCTGCAACCCCGCTGCAAGCCGCGAGCGAGAGCTGAAGAGCTGTTCAAC	99.8%	1475.8	16	2742
241	ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGGTCGCGACACCAAGAGGCC	99.8%	1475.8	16	2742
247	ACCGTGGCCACCTGTACTGCGTGCACGAGAAGATCGAGGTCGCGACACCAAGAGGCC	99.8%	1475.8	16	2742
301	CTGGAACAAGATCGAGAGAGAGAGCAACAAGTCCAGAGAGATCCAGAGGCCGAGGCC	99.8%	1475.8	16	2742
307	CTGGAACAAGATCGAGAGAGAGAGCAACAAGTCCAGAGAGATCCAGAGGCCGAGGCC	99.8%	1475.8	16	2742
361	GCCGACAAGGGCAAGGTGAGCCGAGAACTACCCCATCGTGCAGAACTGCAGGGCCAGATG	99.8%	1475.8	16	2742
367	GCCGACAAGGGCAAGGTGAGCCGAGAACTACCCCATCGTGCAGAACTGCAGGGCCAGATG	99.8%	1475.8	16	2742
421	GTCGACCAAGGCGCATCAGCCCCCGGCAACCTGAAACGCTGGGTGAAGGTGATCGAGAGAG	99.8%	1475.8	16	2742
427	GTCGACCAAGGCGCATCAGCCCCCGGCAACCTGAAACGCTGGGTGAAGGTGATCGAGAGAG	99.8%	1475.8	16	2742
481	GCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGACGAGGCGGCGCAACCCCCAG	99.8%	1475.8	16	2742
487	GCTTCAGCCCCGAGGTGATCCCATGTTCAACCGCCCTGACGAGGCGGCGCAACCCCCAG	99.8%	1475.8	16	2742
541	GACCTGAAACAGATGTTGAACAACCGTGGGCGGCAACGAGGCCCATGCAAGTCTGAAG	99.8%	1475.8	16	2742
547	GACCTGAAACAGATGTTGAACAACCGTGGGCGGCAACGAGGCCCATGCAAGTCTGAAG	99.8%	1475.8	16	2742
601	GACACCATCAACGAGGAGGCGCGAGTGGGACCGCGTGCAACCCCGTGACCGCGGCCCC	99.8%	1475.8	16	2742
607	GACACCATCAACGAGGAGGCGCGAGTGGGACCGCGTGCAACCCCGTGACCGCGGCCCC	99.8%	1475.8	16	2742
661	ATCGCCCCCGGCGAGATGCGCGGAGCCCCGCGGCAAGCAATCGCCGCGACCAAGACACC	99.8%	1475.8	16	2742
667	ATCGCCCCCGGCGAGATGCGCGGAGCCCCGCGGCAAGCAATCGCCGCGACCAAGACACC	99.8%	1475.8	16	2742
721	CTGAGAGAGAGATCGCTGGAAGCAACCCCGCATCCCGTGGGCGCAATCTTAC	99.8%	1475.8	16	2742
727	CTGAGAGAGAGATCGCTGGAAGCAACCCCGCATCCCGTGGGCGCAATCTTAC	99.8%	1475.8	16	2742
781	AAGCGGTGATCATCTCGGCGCTGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC	99.8%	1475.8	16	2742
787	AAGCGGTGATCATCTCGGCGCTGAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC	99.8%	1475.8	16	2742
841	CTGACATCAAGCAAGGCGCCCAAGAGCCCTTCGCGCACTACGTGACCGCTTCTTCAAG	99.8%	1475.8	16	2742
847	CTGACATCAAGCAAGGCGCCCAAGAGCCCTTCGCGCACTACGTGACCGCTTCTTCAAG	99.8%	1475.8	16	2742

QY	901	ACCCTGCGCGCCGAGCAGAGCACCCAGAGGTTGAAGAACTGATGAACCGACACCTGCTG	960
Db	907	ACCCTGCGCGCCGAGCAGAGCACCCAGAGGTTGAAGAACTGATGAACCGACACCTGCTG	966
QY	961	GTGCAGAACGCCAACCCCGACTGCAAGAACCATCTGCGGCTCTCGGCCCGCGCCAGC	1020
Db	967	GTGCAGAACGCCAACCCCGACTGCAAGAACCATCTGCGGCTCTCGGCCCGCGCCAGC	1026
QY	1021	CTGAGAGAGATGATGACCGCGCTGCCAGGGCGTGGGCGGCCCAAGCACAAGGCCGCGTG	1080
Db	1027	CTGAGAGAGATGATGACCGCGCTGCCAGGGCGTGGGCGGCCCAAGCACAAGGCCGCGTG	1086
QY	1081	CTGGCCGAGGCGATGAGCCAGGCCAACACCAAGCGTGATGATGACAGAAGCAACTTCAAG	1140
Db	1087	CTGGCCGAGGCGATGAGCCAGGCCAACACCAAGCGTGATGATGACAGAAGCAACTTCAAG	1146
QY	1141	GGCCCCCGCGCGCATGCTCAAGTGCTTCAACTGGGCAAGAGGGCCACATCGCCCCGAAC	1200
Db	1147	GGCCCCCGCGCGCATGCTCAAGTGCTTCAACTGGGCAAGAGGGCCACATCGCCCCGAAC	1206
QY	1201	TGCGCGCGCCCCCGCAAGAGGGCTGCTGAAGTGCAGAGGAGGGCCACCATGTAAG	1260
Db	1207	TGCGCGCGCCCCCGCAAGAGGGCTGCTGAAGTGCAGAGGAGGGCCACCATGTAAG	1266
QY	1261	GACTGCACCGAGCGCCAGGCCCACTTCTGGGCAAGATCTGGGCCAGCCACAAGGGCCG	1320
Db	1267	GACTGCACCGAGCGCCAGGCCCACTTCTGGGCAAGATCTGGGCCAGCCACAAGGGCCG	1326
QY	1321	CCCGGCAACTTCTCTGCAGAGCGCGCCCGAGGCCACCGCCCCCCCCCGGAGACTTCCGC	1380
Db	1327	CCCGGCAACTTCTCTGCAGAGCGCGCCCGAGGCCACCGCCCCCCCCCGGAGACTTCCGC	1386
QY	1381	TTTCGAGGAGACCAACCCCCGGGCCAAGACAGAGACCAAGGACCGCGAGACCTTGACCAAC	1440
Db	1387	TTTCGAGGAGACCAACCCCCGGGCCAAGACAGAGACCAAGGACCGCGAGACCTTGACCAAC	1446
QY	1441	CTGAAGAGCCTGTTCGGCAACGACCCCTTGAGCCACTAA	1479
Db	1447	CTGAAGAGCCTGTTCGGCAACGACCCCTTGAGCCACTAA	1485

```

RESULT 15
US-10-190-435-9
; Sequence 9, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmut_C
US-10-190-435-9

```

Query Match	99.8%	Score 1475.8;	DB 15;	Length 3930;
Best local Similarity	99.9%;	Pred. No. 1.9e-307;		
Matches 1477; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	ATGGGGCGCCGCGCAGCATCTCTGCGCGCGGCGAAGTGTGACCGCTTGGAGCGCATCCGC	60	
Db	7	ATGGGGCGCCGCGCAGCATCTCTGCGCGCGGCGAAGTGTGACCGCTTGGAGCGCATCCGC	65	

QY	61	CTGGCCCCCGCGCAAGAGTGTCA	CATGATGAAGCACTGGTGTGGCCAGCCGCGAG	120			
Db	67	CTGGCCCCCGCGCAAGAGTGTCTA	CATGATGAAGCACTGGTGTGGCCAGCCGCGAG	126			
QY	121	CTGGAGAAGTTGCGCCCTGAA	CCCCCGGCTGTGGAGACCAGGAGGGCTGCAGCAGATC	180			
Db	127	CTGGAGAAGTTGCGCCCTGAA	CCCCCGGCTGTGGAGACCAGGAGGGCTGCAGCAGATC	186			
QY	181	ATCCGCCAGCTGCA	CCCCGCTTCGAGACCGGCA	GGCAGGAGACTGAAGACCTTCTCAAC	240		
Db	187	ATCCGCCAGCTGCA	CCCCGCTTCGAGACCGGCA	GGCAGGAGACTGAAGACCTTCTCAAC	246		
QY	241	ACCGTGGCCACCCCTGTACT	GCCTGTCACGAGAAAGATCGAGGTTC	CGGACACCAAGAGGCC	300		
Db	247	ACCGTGGCCACCCCTGTACT	GCCTGTCACGAGAAAGATCGAGGTTC	CGGACACCAAGAGGCC	306		
QY	301	CTGGACAAGATCGAGGAGGAG	AGCAACAAGTGTCCAGCAGAAGATCCAGCAGGCCGAGGCC	360			
Db	307	CTGGACAAGATCGAGGAGGAG	AGCAACAAGTGTCCAGCAGAAGATCCAGCAGGCCGAGGCC	366			
QY	361	GCCGACAAGGGCAAGGTGAG	CCCACTA	CCCCCATCTGTGCAGAACTTCGAGGGCCAGATG	420		
Db	367	GCCGACAAGGGCAAGGTGAG	CCCACTA	CCCCCATCTGTGCAGAACTTCGAGGGCCAGATG	426		
QY	421	GTGCAACCAAGGCCATCA	GGCCCCCGCA	CCCTGAA	CGCCTGGGTGAAGGTGATCGAGGAGAAG	480	
Db	427	GTGCAACCAAGGCCATCA	GGCCCCCGCA	CCCTGAA	CGCCTGGGTGAAGGTGATCGAGGAGAAG	486	
QY	481	GCCTTCAGCCCCCGAGGTGAT	CCCATGTTCA	CCGCCCTGA	CGGAGGGCGCCACCCCCCAG	540	
Db	487	GCCTTCAGCCCCCGAGGTGAT	CCCATGTTCA	CCGCCCTGA	CGGAGGGCGCCACCCCCCAG	546	
QY	541	GACCTGAACA	CGATGTTGAACA	CCGTGGCGGCCACCA	GGCCGCATGCA	GATGCTGAAG	600
Db	547	GACCTGAACA	CGATGTTGAACA	CCGTGGCGGCCACCA	GGCCGCATGCA	GATGCTGAAG	606
QY	601	GACACCATCA	ACGAGGAGGCGCCGAGTGG	GA	CCCGCTGCA	CCCCCTGCACGCGCGCCCC	660
Db	607	GACACCATCA	ACGAGGAGGCGCCGAGTGG	GA	CCCGCTGCA	CCCCCTGCACGCGCGCCCC	666
QY	661	ATGCCCCCGGCGCAGATG	CGCGAGCCCCCGGCGCAGCA	CA	TCGCGCGGCA	CCAGCACC	720
Db	667	ATGCCCCCGGCGCAGATG	CGCGAGCCCCCGGCGCAGCA	CA	TCGCGCGGCA	CCAGCACC	726
QY	721	CTGCAGGAGCAGATCG	CTGGATGACCA	CCCCCCCAT	CCCCCGTGGCGCATCTAC		780
Db	727	CTGCAGGAGCAGATCG	CTGGATGACCA	CCCCCCCAT	CCCCCGTGGCGCATCTAC		786
QY	781	AAGCGGTGATCAT	CTGGGCGCTGAACA	CAAGATCGTGC	GATGTACAG	CCCCGTGAGCATC	840
Db	787	AAGCGGTGATCAT	CTGGGCGCTGAACA	CAAGATCGTGC	GATGTACAG	CCCCGTGAGCATC	846
QY	841	CTGCACATCA	ACGAGGCGCCCAAGAG	CCCTTCGCGCACTA	CGTGAAC	CGCTTCTTCAAG	900
Db	847	CTGCACATCA	ACGAGGCGCCCAAGAG	CCCTTCGCGCACTA	CGTGAAC	CGCTTCTTCAAG	906
QY	901	ACCTTGCGCGCGAGCAG	AGCA	CCAGAGGTGA	AGAACTGATGAC	CGACACCCCTGCTG	960
Db	907	ACCTTGCGCGCGAGCAG	AGCA	CCAGAGGTGA	AGAACTGATGAC	CGACACCCCTGCTG	966
QY	961	GTGCAGAACGCCAA	CCCCCGACTGCA	AGA	CCATCTGCGCGCTCTCG	CGCCCGGCGCACG	1020
Db	967	GTGCAGAACGCCAA	CCCCCGACTGCA	AGA	CCATCTGCGCGCTCTCG	CGCCCGGCGCACG	1026
QY	1021	CTGGAGGAGATGATGAC	CGCCTG	CCAGGGCGTGGGCGG	CCAGCACAAG	CGCCGCTG	1080
Db	1027	CTGGAGGAGATGATGAC	CGCCTG	CCAGGGCGTGGGCGG	CCAGCACAAG	CGCCGCTG	1086
QY	1081	CTGGCCGAGCGGATGAG	CCAGGCCAA	CAACCA	CGTGTATGTGCA	AGAGCAACTTCAAG	1140
Db	1087	CTGGCCGAGCGGATGAG	CCAGGCCAA	CAACCA	CGTGTATGTGCA	AGAGCAACTTCAAG	1146

QY	1141	GGCCCCCGGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCGCAAC	1200
Db	1147	GGCCCCCGGCGCATCGTCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCGCAAC	1206
QY	1201	TGCGCGGCCCCCGGCAAGAGGGGTGTGGAAGTGTGGGCAAGAGGGCCACCAAGATGAAG	1260
Db	1207	TGCGCGGCCCCCGGCAAGAGGGGTGTGGAAGTGTGGGCAAGAGGGCCACCAAGATGAAG	1266
QY	1261	GACTGCACCGAGCGCGCAAGGCTTCTGGGCAAGATCTGGCCCAAGCCCAAGGGCGGC	1320
Db	1267	GACTGCACCGAGCGCGCAAGGCTTCTGGGCAAGATCTGGCCCAAGCCCAAGGGCGGC	1326
QY	1321	CCCGGCAACTTCTTGAGAGCGCGCCCGAGCCCAACCGCCCCCGCGGAGAGCTTCGC	1380
Db	1327	CCCGGCAACTTCTTGAGAGCGCGCCCGAGCCCAACCGCCCCCGCGGAGAGCTTCGC	1386
QY	1381	TTGAGGAGACCAACCCCGGCGCAAGAGAGAGCAAGCAACCGCGAGACCTTGACCAAC	1440
Db	1387	TTGAGGAGACCAACCCCGGCGCAAGAGAGAGCAAGCAACCGCGAGACCTTGACCAAC	1446
QY	1441	CTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAGTAA	1479
Db	1447	CTGAAGAGCCTGTTCGGCAACGACCCCTGAGCCAGTAA	1485

Search completed: March 13, 2005, 04:29:36  
 Job time : 901.894 secs

THIS PAGE BLANK (USPTO)